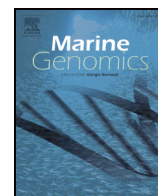




Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Review

Modelling plankton ecosystems in the meta-omics era. Are we ready?

Krzysztof Franciszek Stec^{a,1}, Luigi Caputi^{a,1}, Pier Luigi Buttigieg^{b,1}, Domenico D'Alelio^a,
 Federico Matias Ibarbalz^c, Matthew B. Sullivan^d, Samuel Chaffron^e, Chris Bowler^c,
 Maurizio Ribera d'Alcalà^{a,*}, Daniele Iudicone^{a,*}

^a Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy^b Alfred-Wegener Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany^c Ecole Normale Supérieure, PSL Research University, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR 8197, INSERM U1024, 46 rue d'Ulm, F-75005 Paris, France^d Ohio State Department of Microbiology, The Ohio State University, Columbus, United States^e Laboratoire des Sciences du Numérique de Nantes (LS2N), UMR 6004 – Université de Nantes, CNRS, École Centrale de Nantes, IMT Atlantique, 2 rue de la Houssinière, 44322 Nantes, France

ARTICLE INFO

Article history:

Received 20 November 2016

Received in revised form 24 February 2017

Accepted 25 February 2017

Available online xxx

ABSTRACT

Recent progress in applying meta-omics approaches to the study of marine ecosystems potentially allows scientists to study the genetic and functional diversity of plankton at an unprecedented depth and with enhanced precision. However, while a range of persistent technical issues still need to be resolved, a much greater obstacle currently preventing a complete and integrated view of the marine ecosystem is the absence of a clear conceptual framework. Herein, we discuss the knowledge that has thus far been derived from conceptual and statistical modelling of marine plankton ecosystems, and illustrate the potential power of integrated meta-omics approaches in the field. We then propose the use of a semantic framework is necessary to support integrative ecological modelling in the meta-omics era, particularly when having to face the increased interdisciplinarity needed to address global issues related to climate change.

© 2017 Published by Elsevier B.V.

Contents

1. Introduction	0
2. Core challenges of modelling plankton in marine ecosystems	0
3. Part I: the distinctiveness of plankton and their habitat	0
3.1. Different models of plankton diversity, competition and their impact on biogeochemical fluxes before omics	0
3.2. Mechanistic physiological models and the need to reduce the degrees of freedom	0
3.3. Dealing with complexity	0
3.4. Statistical modelling: a bridge between empirical and model-based understanding	0
3.5. Inclusion of evolutionary processes	0
3.6. A preliminary synthesis.	0
4. Part II: meta-omics approaches to biodiversity and functional diversity modelling of the plankton	0
4.1. From genes to trophic webs. Are we ready?	0
4.2. Integrated meta-omics approach to plankton biodiversity and functional diversity	0
4.3. The neglected marine virome	0
4.4. From meta-omics to modelling	0
5. Part III. Semantic approaches in support of conceptual modelling	0
6. Conclusions	0
Acknowledgements	0
References	0

* Corresponding authors.

E-mail addresses: maurizio@szn.it (M. Ribera d'Alcalà), iudicone@szn.it (D. Iudicone).¹ These authors equally contributed to the paper.<http://dx.doi.org/10.1016/j.margen.2017.02.006>

1874-7787/© 2017 Published by Elsevier B.V.

Please cite this article as: Stec, K.F., et al., Modelling plankton ecosystems in the meta-omics era. Are we ready?, Mar. Genomics (2017), <http://dx.doi.org/10.1016/j.margen.2017.02.006>

1. Introduction

It is very likely that the ongoing wave of dramatic biodiversity loss, which seems to be comparable to the mass extinctions that occurred in the past, is being driven by human impact on the biosphere (Barnosky et al., 2011, Secretariat of the Convention on Biological Diversity). The scale of anthropogenic environmental perturbations has urged both the public and decision-makers to call for tools and approaches allowing the forecasting, management and mitigation of the continuing loss of biodiversity (Barnosky et al., 2011). Such a call can only be met with an integrated approach to understanding the mechanisms that regulate biodiversity, thus linking abiotic and biotic factors within conceptual models of ecosystems (Thuiller et al., 2013).

Since its inception, the study of ecology has focused on the interactions between organisms and between organisms and their environments (Haeckel, 1866). In any given ecosystem, populations of organisms compete and cooperate, forming ecological guilds which are traditionally characterized through metrics that estimate their phylogenetic and functional diversity alongside their response to disturbance. If collected with sufficient spatial and temporal resolution, these metrics permit a description of the way in which a given ecosystem follows a specific trajectory in the space of variables, or phase space, for example that of species abundance. Each trajectory would depend on the rates at which each species responds to the vast array of interactions and fluxes in and out of the system. Such a trajectory, although seldom mentioned as such, is what determines and is behind the more familiar term of 'ecosystem functioning'. The above implies that variation in biodiversity, which would alter the spectrum of biological units and consequently their interactions and functions, will naturally affect broader-scale ecosystem functioning (Krause et al., 2014; Tett et al., 2013; Crise et al., 2015). At times, the processes modulated by organisms in a given system are of interest, in which cases a functional diversity (FD) approach is adopted to understand ecosystem dynamics (Fig. 1a). The FD within an ecosystem would then be the assembly of functional traits of all the constituent organisms (Carmona et al., 2016) and summarize the community's metabolic or phenotypic capacities. However, most of these metrics only offer a loose approximation of the overall configuration of interactions within a community. Currently, a plethora of metrics is used to infer FD (for a review, see Mouchet et al., 2010), each of which has a different ability to describe the status of a system, depending on various aspects (e.g., if a process is dominated by disparate species, metrics emphasizing species richness or divergence may better predict functions affecting production than metrics emphasizing species evenness – Clark et al., 2012). Perhaps the main challenge in correctly estimating FD is the sheer complexity of many ecosystems. This is especially true in species-rich ecosystems, where accounting for and characterizing the role of all species is an immense and often unrealistic task. Compounding this challenge, many traits belong to uncharacterized or poorly understood species. Methods for estimating probable trait values for uncharacterized species have been proposed, based on statistical approaches (Madin et al., 2016; Májeková et al., 2016); however the quantification of the extent to which a given biological unit is performing one or more of its ecological functions at a specific time and in a specific environmental setting is hard to achieve. Nevertheless, in the past decade, research on ecosystem functioning has experienced a shift from a taxonomic to a functional perspective on its participant biotic components (Aerts and Honnay, 2011; Tett et al., 2013; Sutherland et al., 2013 – Fig. 1b).

This shift should be properly interpreted. The taxonomic approach, largely grounded in an evolutionary view, has assumed that by knowing the players it would be possible to determine their interactions and roles in an ecological network. In support of this view is the awareness that units of evolutionary selection are typically seen as organisms or ecological populations. However, this view is weakened by the fact that taxonomic identification, even at high resolution, does not necessarily reveal the functional spectrum of a given organism. The so-called

functional trait-based approach tries to fill this knowledge gap by linking phenotypic potential to organisms, populations, or entire communities regardless of their taxonomic identity. This approach should allow a more robust and process-oriented understanding of why community composition changes under the impact of perturbations and possibly predicts what will happen in future scenarios, including those driven by global climate change (Edwards et al., 2013).

2. Core challenges of modelling plankton in marine ecosystems

The ocean, which covers 72% of the Earth's surface, is believed to be responsible for ~46% of global primary production (Field et al., 1998). Recent estimates suggest a total photic-zone eukaryotic plankton richness of ~150,000 OTUs (De Vargas et al., 2015). Within this vast diversity of organisms primary production is carried out by autotrophic plankton and processed by a myriad of other planktonic organisms before being ultimately conveyed to larger metazoans or exported to the deep layers or to the ocean floor. The diversity of plankton and their internal dynamics are therefore key drivers of the functioning of marine ecosystems. In recent years, technical advances from various fields have been applied to the ocean sciences, broadening the range and the accuracy of measurements and allowing unprecedented interdisciplinary collaboration. An exceptional effort is now required to integrate the astonishing amount and variety of data available into a common framework to reduce gaps in our knowledge of plankton functional biology and ecology.

Novel methodological and technical developments, as well as efforts in cross-disciplinary community building, are enabling for the first time the global and integrated measurement of variables required to assess both biodiversity and FD. The relatively recent applications of sequencing techniques to biodiversity assays, combined with standardized sampling approaches, are revising our view of marine biodiversity (Leray and Knowlton, 2015; Kopf et al., 2015). DNA metabarcoding (e.g., Taberlet et al., 2012; Valentini et al., 2016) is allowing, en masse, species identification from environmental samples, but is often unable to link phylotypes to phenotypes; therefore, taxonomic and functional diversity remain decoupled. The above efforts call for advanced informatics, data mining and analysis, and integrative modelling centered on a common conceptualization and terminology. Modelling in particular, if intended in its wide meaning of reproducing and testing our perception of reality, is likely to play a crucial role as it stimulates hypothetical mechanistic reconstructions of observed dynamics providing the first support or the falsification of working hypotheses.

Despite the endeavors to define and quantify biological traits, extant plankton models typically explore the distribution and diversity patterns of micro-organisms (e.g., Follows et al., 2007) and have seldom exploited trait information. Notably, traits that are unrelated to trophic or defensive requirements, e.g., those controlling differences in plankton life strategies (Anderson, 2005), receive even less attention. These constraints result in a considerable under-representation of both organismal diversity (OD) and FD, and their impacts, in ecosystem models. The goal of the present review is to draft the state-of-the-art in both plankton ecological modelling and ocean meta-omics to pave the way to possible future directions for the integration of meta-omics data in a new generation of models.

To achieve this goal, we first discuss the key, though elusive, processes in plankton dynamics. In particular, we explore processes determining community structure and the short- and long-term succession of species. Secondly, we discuss the advantages and disadvantages of various numerical and statistical models. We then describe and evaluate novelties in meta-omics approaches to marine biological, functional and trophic diversity, highlighting their potential for being incorporated into models and for advancing our knowledge. We then explore the role of semantic technologies in promoting harmonized knowledge – and thus a basis for more interdisciplinary models – across the multiple often disparate domains investigating planktonic systems. Finally, we

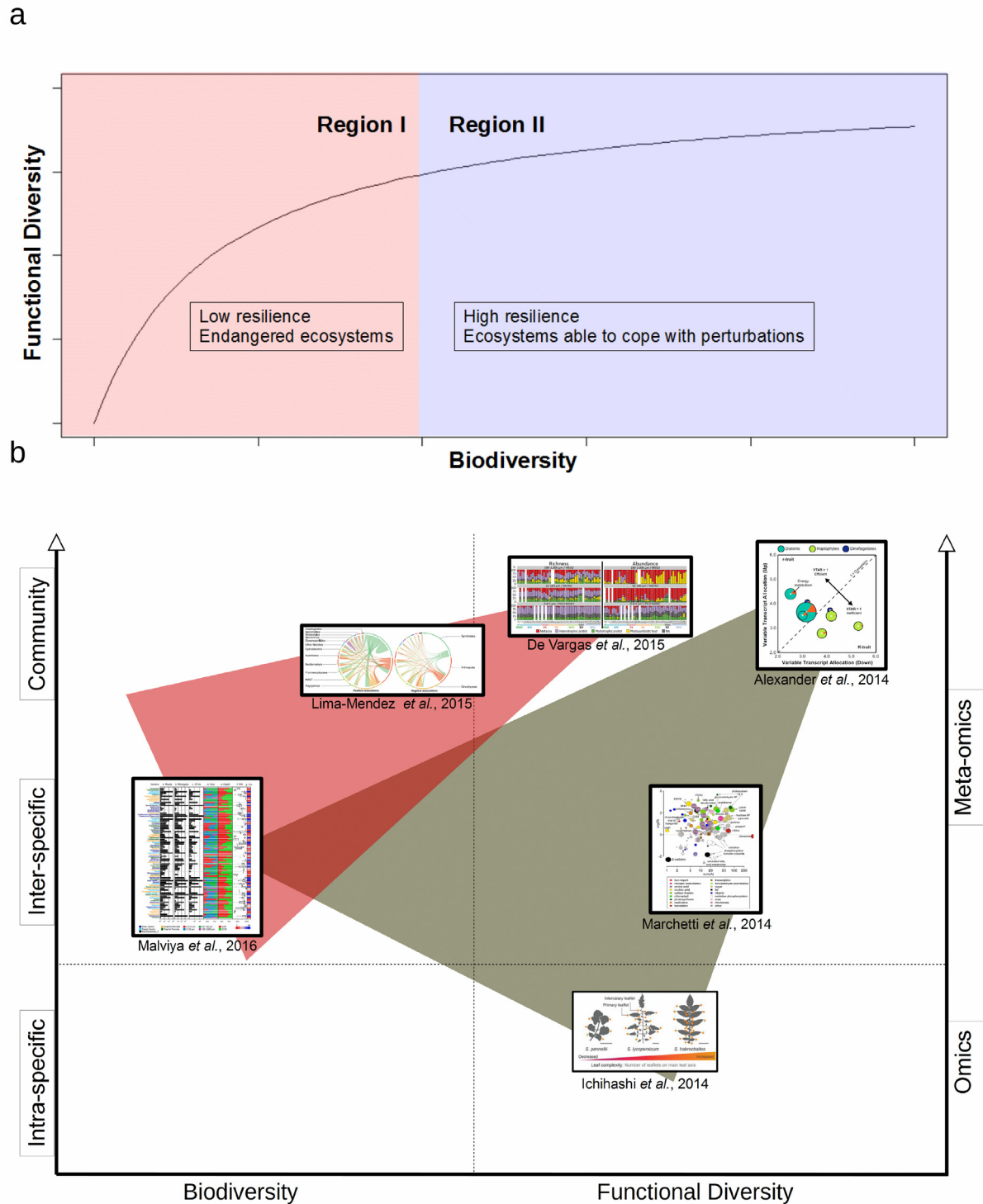


Fig. 1. a) Relationship between biodiversity (x axes) and functional diversity (FD) (y axes). In Region I, FD increases with biodiversity (functions are not redundant) in the ecosystem. After a given threshold, however, functions become redundant (i.e., FD does not increase to any increase in biodiversity (Region II)). Redundancy of functions may help the stability of the system. Any ecosystem placed in Region I is more prone to the effects of disturbance from perturbations, since the gap caused by loss or the change in a function can be hardly filled, while redundancy of functions helps the ecosystems placed in Region II to cope with perturbations (modified after Naeem et al., 1998). b) The impact of meta-omics analyses in deciphering biodiversity and FD. In within-species studies, FD is meant to unveil the individual variability of specific traits which may be important for the ecosystem functioning (i.e., mating preference, stress response, nutrients uptake and photosynthesis). Transcriptional and genomic approaches are thus fundamental to investigate the complexity of the inter-individual variance of traits of interest, as the complexity of leaf-morphology in plants (Ichihashi et al., 2014). In inter-specific studies (i.e., species belonging to a single genus, order of phylum), both biodiversity (red triangle) and FD are important. In barcoding-based studies is possible to get insight in the distribution, even at global scale, and into the phylogenetic relationships within a given taxonomic level (Malviya et al., 2016), while by omics based studies is possible to infer specific traits of interest (Marchetti et al., 2012). Barcoding analyses may be universal (meta-barcoding), i.e. not constrained to a specific group of interest (Lima-Mendez et al., 2015) and are also used, to some extent, to infer FD (De Vargas et al., 2015). However, a much more profound knowledge of FD may be obtained studying specific traits of interest (Alexander et al., 2015) by meta-omics (metatranscriptomics and metagenomics) approaches. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

emphasize the necessity of exposing and standardizing experimental procedures and the importance of accurate reference databases, and draft a preliminary synthesis.

3. Part I: the distinctiveness of plankton and their habitat

The processes driving species aggregation, alternation, and succession in planktonic communities are a recurrent theme of discussion in aquatic ecology (e.g., Hutchinson, 1961; Margalef, 1978; Follows and Dutkiewicz, 2011; Smetacek, 2012; Sauterey et al., 2015). Part of this recurrence is attributable to the continuing uncertainties emerging from limited knowledge about actual plankton diversity and biology. This view is directly reflected in most of the models simulating diversity and succession of phytoplankton. For example, the physiological processes generally factored in to such models are species-specific uptake or utilization rates of a handful of resources (light, macro- and micro-nutrients). These are often combined with factors such as water motion (Hutchinson, 1957; Margalef, 1974) to ultimately determine the gross growth rate of the species. The net growth rate of each species, and therefore its weight in the community, is then derived by considering a mortality term, including but not limited to grazing (Litchman and Klausmeier, 2008; Anderson, 1998). We have only empirical information about most of these processes, for example, nutrient uptake rates in models are still largely based on the scheme introduced by Droop in the 1970s (Droop, 1974). More up-to-date measures are typically derived from *in situ* data, occasionally processed by statistical methods (see below).

Aside from the immediate, *in situ* processes unfolding around and within a planktonic community, the life history of each constituent species strongly affects the dynamics of its population (e.g., D'Alelio et al., 2010) and, often, the larger community. The traits that are tied to an organism's life history have been selected over evolutionary time and determine its current ecological fitness (Flatt and Heyland, 2011). Many phytoplankton species have complex, heteromorphic life cycles (Von Dassow et al., 2014). Life-cycle phase alternations and life-history traits drive species dynamics along more sophisticated paths than those only regulated by proximate factors. We still lack sufficient empirical data about species-specific stage transitions to be able to effectively integrate them into models, in spite of the insights that 'classical', purely biogeochemical, approaches are providing. Thus, further studies elucidating the role of conserved traits are urgently needed to provide a better understanding of plankton biology and ecology.

An improved understanding of plankton biology and ecology must, at its core, factor in the complexities of plankton habitats. Plankton are drifters, and their evolution and ecology should be analyzed in the context of both global currents and within the mesoscale/local dynamics of vertical and horizontal mixing. The ocean cannot be considered as a uniform mass of water: it is an environment of spatial (e.g., latitudinal, longitudinal, depth-based and coastally determined) and temporal (e.g., annual, seasonal, diel) gradients of temperature, light, salinity, gases, pH, macro- and micro- nutrients, and their interactions.

Further, currents connect distant regions of the ocean and can potentially transfer any planktonic organism from one site to another with the only constraints being the Lagrangian distance, i.e., the length of the path followed by the flow and the time spent to reach the site, and the processes occurring along that path. In other words, in the ocean there are no absolute physical barriers to horizontal dispersal but there are only dynamical barriers such as fronts or environmental gradients. This raises a series of problems concerning the effectiveness of sampling procedures designed for biogeographic purposes. Indeed, the apparent absence of well-defined constraints to dispersal in marine pelagic systems coupled with the evidence of biogeographical patterns (e.g., Reygondeau et al., 2013) has led to assertions about organismal dispersion such as "everything is everywhere, but the environment selects" (Baas Becking, 1934). This assertion, originally intended only for microbes, suggests that local environmental properties strongly

modulate the abundance of components which would, otherwise, be homogeneous. Various investigations, however, have shown this assertion to be untrue, even for marine bacteria (Valentine et al., 2013; Sul et al., 2013; Sauterey et al., 2015), meaning that such marine organisms do have biogeographies that can be explored. Nonetheless, methodological problems still remain. The physical movements of water masses make samples taken from close locations, or taken in the same location over a short span of time (Seymour et al., 2012; Needham et al., 2013), very different. On the other hand, at large scales, surface water currents may constrain community dispersal by implementing physical boundaries (Wilkins et al., 2012). Furthermore, it is now emerging that characterization of the biogeography of marine plankton could alternatively be a biogeography of marine plankton traits (see below). Indeed, focusing on the responses of functional traits and how these determine the relative fitness of a planktonic organism along environmental and biological gradients can be far more informative than taxonomic distribution when attempting to grasp the properties of ecological communities (Barton et al., 2013). A corollary of such studies is that the mere sampling of organisms does not allow for the reconstruction of the driving mechanisms unless environmental and physiological data are also collected. In this view, the meta-omics approach, when combined with sufficient contextual information, may reveal its power and utility (see part II of the present work).

3.1. Different models of plankton diversity, competition and their impact on biogeochemical fluxes before omics

Mathematical models in the marine sciences have frequently been used to study a broad range of processes including oceanic biogeochemical fluxes and their response to climate change, the development and potential consequences of harmful algal blooms, and end-to-end biomass transfers and their socio-economic impacts. These models vary in terms of their structure, complexity, equations and parameterization, but above all in terms of objectives and questions they address.

Most conceptual (and numerical) models concerning marine plankton assume quasi-linear transduction of matter and energy among organisms, with fluid dynamics as the main modulator on short time scales and elemental ratios being an important, if not the dominant, driver on longer time scales. This approach generally results in the creation of blooms of one or of a very limited number of species. A higher OD can be obtained in the simulations if the external forcing (winds, heat fluxes, etc.) of the virtual system display a high (often stochastic) variability, i.e., if the model mimics a system with an intermediate level of disturbance (e.g. Elliott et al., 2001) or if the model allows for internal oscillations due to nonlinear prey-predator couplings (e.g., Dakos et al., 2009). Theory (Tilman et al., 1997; Yachi and Loreau, 1999) predicts that the stability and functioning of the ecosystem can be influenced by both the composition and the richness of taxa (Cardinale et al., 2012). The hypothesis that diversity and abundance of primary producers influences the functioning of ecosystems in terms of primary production, total biomass, or nutrient use, has received considerable empirical support in recent years (Ptacnik et al., 2008; Cardinale et al., 2011). Phytoplankton diversity, in particular, influences ecological and biogeochemical processes as it relates to community composition as well as the density with which niche space is filled at the foundation of the pelagic food web (Duffy and Stachowicz, 2006). The composition of the phytoplankton community may affect the export of organic matter from the surface to the deep ocean, and thereby the global cycles of nitrogen, phosphorus, and carbon (Redfield, 1958; Falkowski et al., 1998; Caron and Countway, 2009), including uptake of atmospheric CO₂ (e.g., Arrigo et al., 1999; Finkel and Irwin, 2000; Behl et al., 2011) and carbon sedimentation (Sieracki et al., 1993; Assmy et al., 2013). Phytoplankton community structure also affects the diversity, activity, and community structure of zooplankton and heterotrophic bacteria by determining the range of available substrates and particle-associated microenvironments. These basal effects propagate up through the food

web to fisheries and humans as end-users of the fisheries (Richardson and Schoeman, 2004; Ainsworth et al., 2011). Given the inherent complexities described above, how can models help to dissect the multifaceted processes at work in planktonic systems, and propose predictions about future scenarios?

An overwhelming number of predictions on future trends in pelagic marine ecosystems is based on models in the form of 'Nutrient-Phytoplankton-Zooplankton-Detritus' (NPZD) configurations, in which the named components compose the main model structure and all the processes are governed by the fluxes among those compartments (e.g., Steele, 1974; Wroblewski et al., 1988; Kawamiya et al., 2000; Palmer and Totterdell, 2001; Anderson and Pondaven, 2003; Schartau and Oeschlies, 2003; Sauterey et al., 2015). The early models, due to their simplicity and computational limitations, restrict the biogeochemical flows to a single component (frequently C or N; Evans and Parslow, 1985). Currently however there is a tendency to augment model formulation by the representation of multiple elements (Follows et al., 2007; Litchman and Klausmeier, 2008). This approach has been formalized in the concept of ecological stoichiometry (Sturner and Elser, 2002). The intrinsic limits of NPZD models include the lack of significant biological feedbacks, other than the use of resources towards maxima determined by external and/or internal concentrations, and no accounting of life strategies. These limits have stimulated the birth of trait-based approaches.

The importance of a trait-based approach, which means that phenotypic specificities of organisms are taken into account, is that they hold the potential to better explain the organization of ecological communities and predict their reorganizations under global change (Litchman and Klausmeier, 2008; Litchman et al., 2007; Finkel, 2001; Armstrong, 1999). As we will discuss below, the trait-based approach is also often conceived as a way to keep the number of variables low by focusing on the variance of the essential variables including Functional Group-specific growth rates. It has long been recognized that microbial growth, respiration, nutrient acquisition, light harvesting, and sinking speed scale up with cell volume or mass (e.g., Laws, 1975; Banse, 1982; Tang, 1995; Litchman et al., 2007; Finkel, 2001; Kooijman, 2001). Consequently, cell size is used as a key trait and allows modelers to reduce the region of possible trait space in ecosystem models (Litchman and Klausmeier, 2008). It has been successfully implemented in models simulating relationships between phytoplankton cell size, abundance, respiration, and links to environmental controls (e.g., Laws, 1975; Irwin et al., 2006).

In trait-based approaches, trait variables are weighted according to the trade-offs among them which constrain their ranges; in other words, the costs and benefits of manifesting a particular physiological characteristic are weighted against those of manifesting another (e.g., Margalef, 1978; Reynolds, 1984). The requirement of traits and trade-offs is also of prime significance for phytoplankton functional type (PFT) models, which are conceptually the same as trait-based models, but group organisms with assumed similar trophic and reproductive characteristics or physiology, thus introducing an ecological guild, labeled with a sort of 'operational' taxon, instead of a set of common traits (Le Quere et al., 2005). Current marine ecosystem models, mainly aimed at simulating biogeochemical fluxes, deal with plankton diversity using such a PFT framework (Totterdell et al., 1993; Moore et al., 2002; Anderson, 2005). On the other hand, species aggregated within one PFT might considerably differ in their morphological features, life strategies, and mutual interactions (Anderson, 2005), attributes which are generally neglected in a typical PFT. The need for a better representation of the diversity within PFTs to enable a more accurate representation of phytoplankton distribution patterns has been discussed by Thingstad and Cuevas (2010) and offers a path towards a more comprehensive future for plankton modelling.

PFT-based and trait-based models allow investigation into the mechanisms of phytoplankton coexistence and the consequences of biological diversity on biogeochemical cycling, yet their effective

implementation is challenged by an insufficient amount of field and laboratory measurements, as well as by the numerous limitations mentioned above.

Finally, the persistence and the distribution of species in ecosystems, which is the leading question in many studies, are generally associated with the existence of ecological niches. While variations of this concept exist, niches are often represented as intervals in the hyperspace of biotic and abiotic variables which overlap with the resources and conditions needed by a particular species to persist (Hutchinson, 1957). The major axes in the hyperspace defining ecological niches of phytoplankton are attributes of the physical environment (e.g., temperature, light availability), resources (e.g. carbon and nutrient sources), and the abundance of natural predators *sensu lato* (grazers, parasites, and viruses) (Margalef, 1978; Reynolds, 1984; Tilman et al., 1982). For each of these axes it is possible to identify a whole hierarchy of traits, from the subcellular to the population level, which modulate organismal responses to niche space. These traits can be classified according to their type (e.g., morphological, physiological, behavioural, and those regulating the life cycle) and their ecological function (e.g., reproduction, resource acquisition, and predator avoidance) (Litchman and Klausmeier, 2008). The identification and appropriate ranking of traits, which define the extent of a species' niche, is presently the major hurdle in understanding and modelling the mechanisms by which phytoplankton communities are structured in the environment.

3.2. Mechanistic physiological models and the need to reduce the degrees of freedom

Given the challenges stated above, one path forward is the implementation of mechanistic models and laboratory experiments to determine reasonable trait combinations that correspond to a given niche (Litchman et al., 2007; Finkel, 2001; Armstrong, 1999; Bruggeman and Kooijman, 2007). Several models that explicitly formulate physiological processes have been proposed with the aim of predicting the impact of variable environmental conditions on phytoplankton performance (e.g., Anning et al., 2000) or light-driven succession (Polimene et al., 2013), or simply to test hypotheses about how they may contribute to fitness (e.g., Esposito et al., 2009).

Various attempts have also been made to find unifying, species-independent 'laws' that can reduce the complexity of natural communities. The standard Dynamic Energy Budget (DEB) model (Kooijman, 2010; Sarà et al., 2014) is an example of this. It is metazoan-focused and functional traits-based, and tries to give a mathematical framework for studying higher trophic levels, which could be therefore included in comprehensive modelling approaches. The DEB represents temperature-dependent metabolic processes and is able to predict not only animal growth performances (Saraiva et al., 2012), but also the potential distribution of invasive species (Sarà et al., 2013). Since the DEB is temperature-based, it is also a good candidate for predictions in the context of global-warming.

Another promising approach is the scale-transition theory (STT) (Chesson et al., 2005), which responds to the primary need of integrating modelling and experimental data from individual investigations at local scale in a more general model. The STT predicts the large-scale behaviour of a system on the basis of the interaction of local nonlinear population dynamics and spatial variation of population abundance and/or environmental conditions (Chesson et al., 2005). While this and the other approaches noted in this section offer attractive ways to effectively simplify models, there is growing interest in addressing complexity, as discussed below.

3.3. Dealing with complexity

Natural systems show collective and emergent properties that cannot be derived by simply observing the specific characteristics of isolated elements that directly affect community behaviour (Proulx et al.,

2005; Tëmkin and Eldredge, 2015). Consequently, analytical approaches that consider all the players have been explored. Among them, graph or network analysis is increasingly being used to unveil collective properties such as ‘robustness,’ or the ability of a network/system to resist changes without modifying its initial configuration (Pocock et al., 2012). From an ecological viewpoint, individuals, species, or species-groups are typically rendered as nodes and some form of relationship between them is captured in the edges of such networks. Networks have been mainly used to display interspecific interactions, including trophic relations (Jordán, 2009), mutualism (Suweis et al., 2013) and parasitism (Laliberté and Tylianakis, 2010). Moreover, since networks are thought to represent highly integrated systems, they also allow identification of potential, indirect interactions (i.e., ‘invisible links’ due to positive or negative feedback) through the examination of network topology.

Network approaches have been applied to study plankton community structure (as the pattern of trophic links) at two states, e.g., eutrophic and oligotrophic, of the same system (D’Alelio et al., 2015, 2016). Association networks have also been inferred to study the structure of natural plankton communities across kingdoms and at global scale (Lima-Mendez et al., 2015).

Ecological network-based models will likely complement biodiversity discovery and biogeochemical studies (Hairston and Hairston, 1993; Thompson et al., 2012). However, integrating network-based estimates of community parameters strongly relies on semantic clarity and coherence in the definition of nodes and edges (e.g., Blüthgen et al., 2008; Vasas and Jordán, 2006). Because they are poorly defined, the resulting estimates can be very difficult to integrate into other analyses (see Part III for further perspectives on the role of semantics). Once network reliability has been assessed, several metrics can be computed to study network topology and, consequently, a set of system properties. For example, in networks exploring trophic interactions the inference of ecosystem functioning by interpreting topological properties in networks is largely more trustworthy if links are weighted based on the extent of potential biomass flows throughout the web, which are significantly mediated by organism characteristics (D’Alelio et al., 2016). The estimates of system properties being generated by network analysis are clearly of interest in enhancing existing models; however, modelers should also be aware of estimates of community responses and dependencies being generated by common statistical approaches used in the ecological domain.

3.4. Statistical modelling: a bridge between empirical and model-based understanding

Datasets that capture the responses of multiple organisms in an ecosystem are frequently evaluated with descriptive statistical methods which, at times, allow the creation of statistical models. High-dimensional, multivariate approaches suited to handle complex ecological data (including multi-omic data sets) are being made increasingly accessible to field and laboratory researchers through software solutions and knowledge resources. (see, e.g., Caporaso et al., 2010; McMurdie and Holmes, 2013; Buttigieg and Ramette, 2014; Cao et al., 2016). With these at hand, researchers who employ well-designed sampling and/or experimental strategies are able to produce sophisticated information outputs. In turn, the broader modelling community is provided with empirically derived estimates of factors such as community composition, turnover, and structuring by environmental factors. Statistical models generated by the ecological community often take the form of a regression analysis, in which the effect of a set of explanatory variables upon a set of response variables is summarized by one or more functions. These relationships may be summarized by linear or more complex functions (e.g., high-order polynomials) and be used predictively. In the context of building numerical models for ecological phenomena such as plankton-plankton or plankton-environment interactions, estimates and predictions from statistical approaches can be fed into

numerical models to align them with field observations. However, there are several important caveats to be considered when combining these approaches.

Perhaps foremost among the risks of integrating statistical and deterministic modelling are improper replication and estimates based on poorly designed sampling or experiments. Statistical estimates and models strongly depend on well-designed replication: the causally independent, repeated application of some set of treatments or conditions to independent sets of experimental subjects. Statistical models built upon studies lacking sufficient replication are likely to produce inaccurate, biased, or unstable estimates and predictions. While more care is being taken in contemporary ecological studies, there is still debate about best practices, which is also percolating into omics and plankton research (e.g., Oksanen, 2001; Cottenie and Meester, 2003; Hurlbert, 2004; Prosser, 2010). Another key risk is the temptation to use overly complex and tailored functions to fit a specific set of unruly ecological data. The root of this temptation is likely the same as that which drives acts like “p-hacking”: a desire to engineer and make much of statistical “significance” in place of realistic and generalizable results (see Nuzzo, 2014 and Ioannidis, 2005 for more). Statistical models built under such influences may easily ‘overfit’ a given response in a misguided quest for impressive statistical values of fit, reducing their generalizability and misleading other modelers. Should these major issues be addressed by more studies, the community will be in a far better position to confront other pressing challenges such as the accurate detection of directionality in causal relationships, joining cross-domain drives to improve statistical outputs (e.g. Ioannidis, 2014).

Despite these caveats, there are many reasons to be optimistic when considering how progress in statistical modelling is bringing ecology closer to numerical models. It is encouraging, therefore, that long-term, omically enabled microbial observatories are emerging in marine systems (e.g., Soltwedel et al., 2013) and are paying special attention to establishing robust fusions of biogeochemical, oceanographic, and biological data analyses for improved statistical and numerical modelling (e.g., Buttigieg and Ramette, 2015; Wulff et al., 2016) and that initiatives such as the Genomic Observatories Network (Davies et al., 2014) are providing frameworks through which to disseminate best practices in data and knowledge integration.

3.5. Inclusion of evolutionary processes

The approaches described above assume time scales with no change in biological traits of plankton. The evolutionary adaptive potential of marine microbial systems under environmental change has also been explored by introducing explicit Darwinian adaptation and simulating evolving phytoplankton communities in space and time (Sauterey et al., 2015). Specifically, tools from adaptive dynamics theory were adopted for evaluating the fitness of invading mutants over annual timescales, and replacing the resident if a fitter mutant arises. Using an evolutionary framework, how community assembly, specifically the emergence of phytoplankton cell size diversity, reflects the combined effects of bottom-up and top-down controls, has been examined.

Empirical studies of community dynamics that explicitly consider evolutionary processes support the view that ecologically important traits of a population can be substantially altered over a period of a few generations (Hendry and Kinnison, 1999; Ellner et al., 2011). Examples of eco-evolutionary feedbacks in nature have emerged as a result of predator-prey (Becks et al., 2012; Yoshida et al., 2003), host-parasite or host-pathogen interactions (Penczykowski et al., 2011). These feedbacks play important roles in community and ecosystem functioning as they co-determine the dynamic behaviour of ecological communities (Becks et al., 2012; Yoshida et al., 2003). Coordinated experimental and modelling studies ought to investigate how pervasive eco-evolutionary feedback loops are for plankton and examine the role that eco-evolutionary processes play in ecosystem functioning.

The optimal traits selected by applying the adaptive model approach are very similar to those selected via the species-selection approach, but the emerging patterns are more robust and the predicted emerging diversity is different (Sauterey et al., 2015). Indeed, as recently demonstrated (Irwin et al., 2015), assuming a lack of adaptive potential in phytoplankton is unrealistic and this assumption should be removed from phytoplankton modelling, although the extent of adaptive potential is currently unknown. Interestingly, while most phytoplankton species may adapt their realized niche to environmental variations in temperature and irradiance, the same is not true for nitrate (Irwin et al., 2015), posing interesting questions about how climate change and global warming may affect phytoplankton. A recent paper (Barton et al., 2016) demonstrated the ‘shift and shuffle’ (i.e., entire communities moving in space - shifting - and communities internally reassembling - shuffling) caused by variations in multiple environmental factors. This work reinforces the idea that meta-data, rather than single variables, are mandatory for evaluating variations in marine communities, and thus for their accurate modelling.

In conclusion, these approaches have demonstrated that explicit evolutionary approaches to model marine microbial populations and functionality are feasible and practical in time-varying, space-resolving settings. These pioneering studies provide new tools for exploring evolutionary interactions on a range of timescales in the ocean.

3.6. A preliminary synthesis

A suite of models aimed at reconstructing plankton diversity, distribution, and impact on Earth system functioning are available. Many of them display some predictive power in terms of fluxes of elements, with a lower capability to predict the patterns of diversity in space and time. In some instances statistical models, which just capture and extrapolate the observed patterns, perform equally well. Besides the lack of a sufficient amount of data about known processes there is a deep gap of knowledge about species-specific functions and species-species interactions. In many modelling approaches, key traits that are common to the large majority of organisms are modelled in a species-independent fashion, such that species disappear and are grouped into compartments (NPZD), or PFTs, or into specific traits (see earlier). These approaches are far from being able to predict if and how diversity will change in the plankton, which species will change their role and if and how this will affect Earth system functioning. This is partly due to the insufficient knowledge about the significance of diversity in organismal biology. To what extent could omics change this scenario?

4. Part II: meta-omics approaches to biodiversity and functional diversity modelling of the plankton

4.1. From genes to trophic webs. Are we ready?

In recent years, the wealth of data coming from global genomic surveys (e.g., GOS and Tara Oceans) has increased significantly. For the first time ever, scientific operators have sufficient data to allow them to predict biologically-based traits from genomics and transcriptomics data in order to introduce them into plankton models (Fig. 2). Although a number of technical and scientific issues still persist, these seem to be less critical than the absence of an organized theoretical framework for exploiting discoveries on basic processes from laboratory experiments using functional genomics (e.g., Falcatore and Bowler, 2002; Kooistra et al., 2007; Coelho et al., 2013; Mock and Kirkham, 2012; Morrissey et al., 2015). Indeed, the lack of unclear experimental design and of well standardized and fully accepted procedures may result in an accumulation of data that is difficult to be used practically in conceptual models. Herein, we list some basic scientific questions that could be preferentially addressed using data from meta-omics surveys.

Models explicitly formulating physiological processes or food-web interactions (Allen and Polimene, 2011) are very seldom, if ever,

based on the new generation of molecular data. Phytoplankton models, either to analyze specific mechanisms or to study competition, ought to be augmented with additional state variables describing molecular and genomic data (e.g., gene expression profiles that determine the physiological state of a population (Dunlop et al., 2009; Motti, 2012; Sardans et al., 2011; Satinsky et al., 2014a, 2014b; Veldhoen et al., 2012; Kostka et al., 2014)). While population-level models (PLMs) are mostly generalist, as their aim is to provide general explanations (Evans et al., 2013), recent technical advances allow the shift towards Individual Based Models (IBMs). Thus, by building upon genomic data, the models could also be enriched with reliable cell-level systems biology models, thus assimilating processes at the intracellular level (e.g., cell physiology and response to stress) with those at the intercellular/inter-organism level (e.g., grazing, competition). This would result in modelling the metabolic state of the population and increase the model's capability to consistently reconstruct the observed phytoplankton biodiversity and community composition in various environments (Armstrong, 2006; Allen and Polimene, 2011; Hellweger et al., 2016). Another advantage of the individual-focused approach is that modelers may incorporate a whole new set of parameters (e.g., phenotypic heterogeneity, between-individual interactions, experimental RNA and protein measurements) into modelling, which was not possible using the PLMs (Hellweger et al., 2016; Daines et al., 2014).

Genomic information can also be used to identify the molecular mechanisms underpinning organism trophic modes (de Vargas et al., 2015; Hilton et al., 2013). In addition to the phototrophic “producers” and heterotrophic “consumers,” other trophic modes have been recognized, e.g., mixotrophy being a combination of phototrophic and heterotrophic metabolism (Estep et al., 1986; Hartmann et al., 2012) in which cells take up organic material from the external environment as small molecules or macromolecules (Stoecker and Gustafson, 2003). Mixotrophy broadens the pool of available resources as the absorption of organic material or the ingestion of a prey may serve as a source of energy and carbon when photosynthesis is limited by light availability or resources are scarce (Tittel et al., 2003). Those trophic switches also affect the physiological characteristics of species, e.g., the maximum specific growth rates and nutrient affinity of mixotrophic dinoflagellates are typically lower with respect to photoautotrophs (Edwards et al., 2012), and their growth and grazing rates are slower than in similar heterotrophic specialists (Jeong et al., 2010). Indeed, modelling studies exploring the role of mixotrophy have elucidated its role in structuring the plankton community and controlling the biomass transfer up the food chain to larger sizes classes (Ward and Follows, 2016).

Symbiotic interactions representing mutualistic relationships where one species lives within or on another species (Decelle et al., 2012), or parasitism where one organism exists in association with another to the detriment of that partner, must also be considered (Guillou et al., 2008; Chambouvet et al., 2011; Bachvaroff et al., 2012). These life strategies may have a profound impact on population dynamics. Likewise specific strategies such as chain formation, aggregation or changes in morphological characteristics can highly impact the coupling with higher trophic levels (Bergkvist et al., 2012). Molecular information has, indeed, shown that such modes are much more relevant than previously thought (e.g., de Vargas et al., 2015) and molecular tools can allow to characterize much better inter- and intra-specific plankton interactions.

4.2. Integrated meta-omics approach to plankton biodiversity and functional diversity

With the development of high-throughput sequencing techniques, metabarcoding is becoming a pivotal tool in biodiversity studies. However, technical and procedural issues still affect the reliability of metabarcoding-based biodiversity surveys. Different molecular markers have been used in metabarcoding essays (16S, 18S, CO1 and others - Cristescu, 2014). The multiplicity of markers may represent a first

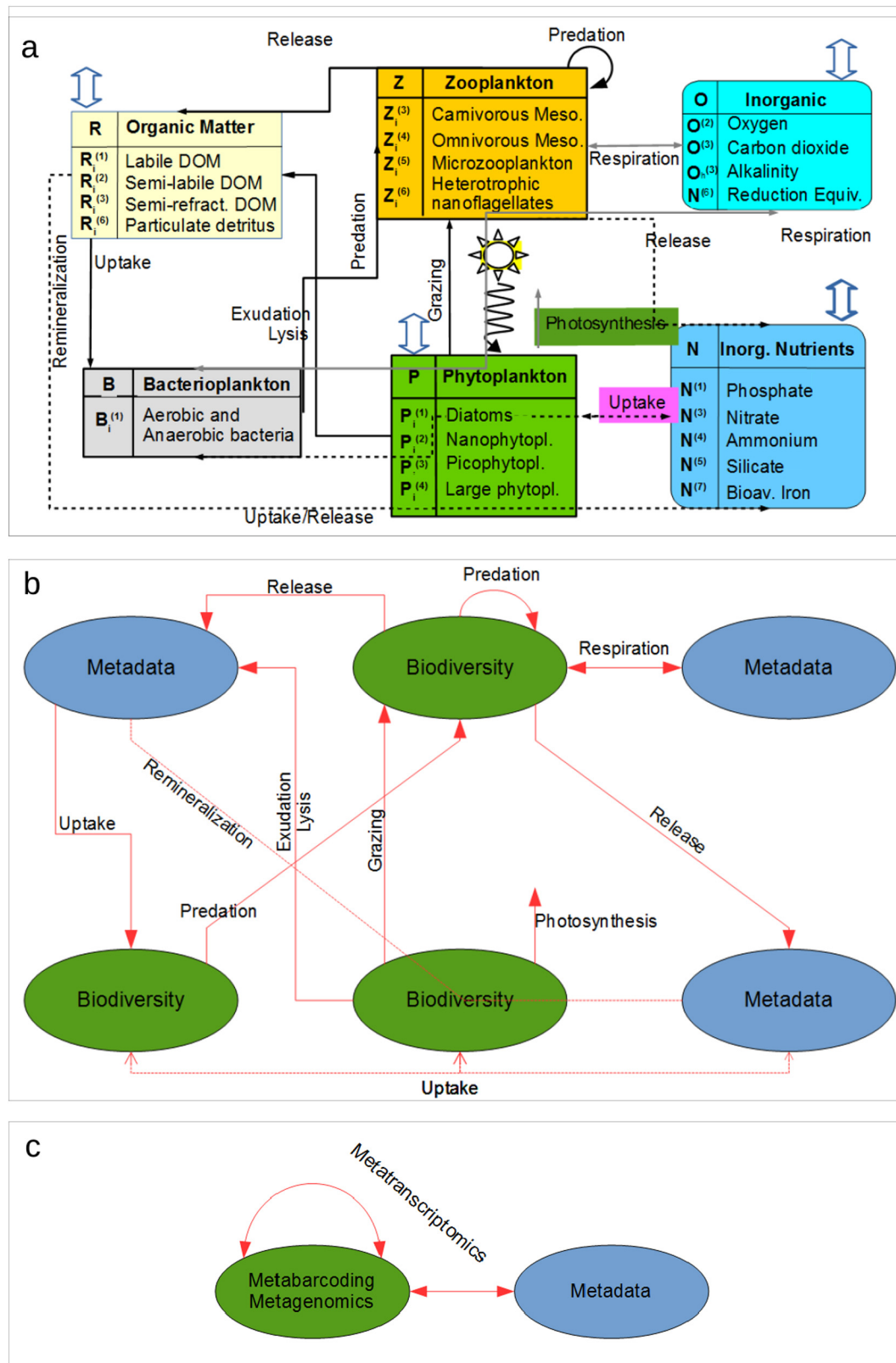


Fig. 2. a) A synthetic scheme of the biogeochemical processes that one of the presently most utilized model, the Biogeochemical Flux Model (BFM) (Vichi et al., 2007), simulates. In light green and magenta are highlighted two processes, photosynthesis and nutrient assimilation. b) A schematic view of how meta-barcoding and meta-omics may be useful to provide the input of the BFM model. Green circles represent input translated from meta-barcoding and meta-genomics (i.e., taxonomic composition, abundance and biodiversity). Red arrows represent processes which are inferred by meta-transcriptomics. c) All the inputs required for the BFM fall into three categories of data: i) diversity (inferred by meta-barcoding and meta-genomics), ii) functional (inferred by meta-transcriptomics) and iii) biogeochemical and ecophysiological meta-data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

downside of metabarcoding because it can be a barrier to survey comparisons. Moreover, the mentioned markers are unfit to uncover the real diversity along the whole evolutionary tree of life (Deagle et al., 2014; Cristescu, 2014; Cowart et al., 2015). Standardization of the molecular marker of choice may represent a significant improvement for the field. In most of the literature, the taxonomic assignment of sequences to species is reference-based. This means that error-free, comprehensive databases are a prerequisite for error-free, reliable metabarcoding studies. However, reference databases are neither error-free nor complete. The incompleteness of databases and sequencing error thus bias our view of metabarcoding-based biodiversity. The scientific community is aware of these problems, and is trying to solve them (Zepeda Mendoza et al., 2014). In this respect, a parallel approach is to build reference-free metabarcoding methods to overcome the problems related to reference database incompleteness and unreliability. However, these methods rely on the clustering of sequences into Operational Taxonomic Units (OTUs) (Blaxter et al., 2005). Unfortunately, the OTU concept as a proxy to species is not widely accepted by the community (Blaxter et al., 2005). To date, OTUs without a taxonomic classification can be only safely used for comparison among sites and, to some extent, for richness estimations, prompting for a more accurate (beyond the mere operational one) definition of OTUs.

Despite these caveats, metabarcoding opens new perspectives in diversity studies. It sets an additional benchmark with respect to classical taxonomic counts because of the much more extensive coverage, both in physical space and in the genetic space occupied by a species.

A significant effort to develop comprehensive and standardized pipelines in metabarcoding studies has been made recently (Schloss et al., 2009; Caporaso et al., 2010; Zimmermann et al., 2014). Today large-scale global marine metabarcoding campaigns enable to fully appreciate the untapped marine biodiversity in the global ocean at an unprecedented level of resolution. In recent publications, de Vargas et al. (2015) and Biard et al. (2016) revealed the enormous importance of Rhizaria in the marine biome, a supergroup which was constantly overlooked in traditional taxonomic identification. Moreover, current models of biodiversity usually rely on an oversimplified view of the trophic relations between species/groups, in which only linear relations between primary producers, grazers and predators are taken into account. This picture is unrealistic. It has been demonstrated that prey-predator relationships are much more complex (Selander et al., 2011). de Vargas et al. (2015) assigned OTUs to trophic categories, enabling the discovery of the astonishing diversity of parasites and symbiotic hosts. Additional studies revealed the astonishing diversity of diplomonads, another heretofore underexplored component of marine plankton ecosystems (David and Archibald, 2016; Flegontova et al., 2016).

While metabarcoding and related phylogenetic analyses are helpful in the assignment of OTUs to a-priori defined functional groups, metagenomics and metatranscriptomics approaches hold the potential to have an even greater impact for characterizing functional diversity. Various publicly accessible databases exist to provide orthologous groups of proteins at different taxonomic levels, each with integrated and summarized functional annotations, two notable examples are the Kyoto Encyclopedia of Genes and Genomes (KEGG – <http://www.genome.jp/kegg/>, Ogata et al., 1999) and the eggNOG 4.5 (<http://eggnogdb.embl.de/#/app/home> – Huerta-Cepas et al., 2015).

Analyses based on meta-omics data allowed to formulate hypotheses about biotic interactions (e.g., Lima-Mendez et al., 2015), to assess the persistence and transport of organisms across basins favored by mesoscale structures (e.g., Villar et al., 2015), to characterize depth partitioning of *Prochlorococcus* populations (e.g., Shi et al., 2011), to study annual patterns of diel and seasonal compositional and functional biodiversity in coastal environments (Gilbert et al., 2010; Gifford et al., 2014), to investigate the microbial community associated with the Amazon River plume (Satinsky et al., 2014a, 2014b), to demonstrate the role of temperature, more than other physical parameters, as the

main driver of the epipelagic community composition (Sunagawa et al., 2015), or to show that taxonomy-based and functional genes-based biogeographic patterns differ greatly (Haggerty and Dinsdale, 2016).

The inference of species associations and community structure from co-occurrence networks using molecular data is becoming routine in microbial ecology (Chaffron et al., 2010; Faust and Raes, 2012; Bálint et al., 2016). These networks are useful to predict species interactions and identify keystone species within a microbial ecosystem, and they can also serve as a framework to study the dynamics of the community structure using time-series data (Fuhrman et al., 2015; Caron et al., 2017). Today, various statistical approaches are available to infer species associations from marker gene abundance data (16S or 18S rRNA gene), but benchmarking these techniques is difficult and they appear to vary widely in sensitivity and precision (Weiss et al., 2016).

An obvious obstacle for training a comprehensive framework from meta-omics approaches is given by the lack of an integrated pipeline allowing the creation of a compendium of quality-controlled and cohesive multi-omics. In this respect, progress has been made in the fields of cellular meta-omics (Kim et al., 2016) and microbial meta-omics (Urich et al., 2014; Franzosa, 2015) of the human gut microbiome. The study of the human gut biome represents a field in which researchers interested in marine meta-omics should look with the utmost interest. Indeed, recent progress to integrate and normalize meta-omics data both at single cell (Kim et al., 2016) and at the whole biome level (Darzi et al., 2015) are allowing the formalization of predictive models, as shown in a recent study on the oral microbiome during in vitro sugar metabolism integrating metatranscriptomic and metabolomic data for analyzing complex ecosystems (Edlund et al., 2015).

In functional diversity studies of marine plankton, acceptable levels of pipeline formalization and normalization have been not yet reached. Yet, the power of an integrated meta-omics approach, coupled with and framed within a theoretical conceptual framework, appears necessary to both fully profit from the advantages of NGS techniques, both to transform huge numerical databases into interpretable data and meaningful concepts.

4.3. The neglected marine virome

An entire component of the global ecosystem has been totally neglected in ecological modelling and conceptualization for years: marine viruses. Viruses are known to be at least as abundant as microbial cells (Wigington et al., 2009) and to impact microbes through mortality, horizontal gene transfer and metabolic reprogramming (Breitbart, 2012; Brum and Sullivan, 2015; Fuhrman, 1999; Suttle, 2007; Weinbauer, 2004). Modelling studies suggest that viruses may affect ecosystem functioning and community structure by increasing organic matter recycling, reducing transfer to higher trophic levels and increasing net primary productivity (Weitz et al., 2015). However, viral ecology has been challenged to extend beyond anecdotal or qualitative measurements because most host cells and therefore viruses cannot be cultured (Edwards and Rohwer, 2005) and cultivation-independent diversity studies are limited by the fact that viruses do not share a single gene that can be used as a barcode, and reference databases are so poor that most viruses observed in nature remain 'unknown' (Brum et al., 2015a, 2015b).

Fortunately, viral communities can now be quantitatively surveyed, at least for double-stranded DNA viruses (Duhaime et al., 2012; Solonenko and Sullivan, 2013). This unknown viral sequence space is being rapidly 'mapped' (Brum et al., 2015a, 2015b; Mizuno et al., 2013; Roux et al., 2015), and analytical approaches have been pioneered to circumvent the lack of barcode genes to enable ecological studies at the level of the gene, population and community (reviewed in Brum and Sullivan, 2015). Methodological variants of community survey techniques also now enable quantification of both lytic and temperate viral life styles (Brum et al., 2016), simultaneous observation of millions

of viruses for a single host (Deng et al., 2014), and the capture of large fragments of viral genomes from microbial genomic and metagenomic datasets (Mizuno et al., 2013; Roux et al., 2015; Labonté et al., 2015).

Together these advances have led to a renaissance in viral ecology, particularly in the oceans, whereby niche-defining and auxiliary metabolic genes, population diversity, and drivers of community structure can now be routinely explored in complex viral communities (reviewed in Brum and Sullivan, 2015; Sullivan, 2015). Further, recent coupling of large-scale quantitative viral metagenomic datasets with paired microbial and physico-chemical datasets, as well as advanced analytical networking and modelling approaches have revealed virus-host interaction networks (Lima-Mendez et al., 2015) and suggested that viruses strongly influence carbon flux in the oceans (Guidi et al., 2016). The integration of viral meta-omics into ecological modelling is challenging, yet it will represent a significant advance for establishing realistic models of marine plankton dynamics. In this framework, it has been demonstrated that viral metagenomics may be used to infer the drivers of viral community structure (Hurwitz et al., 2014). This finding, besides illustrating the power of omics data coupled with advanced bioinformatic tools, may represent important advance for predictively modelling the structure of marine viruses and of their microbial hosts.

4.4. From meta-omics to modelling

Statistical approaches inherited from fields such as numerical ecology (e.g., Legendre and Legendre, 2012), methods that explicitly handle sparsity (e.g., Zou et al., 2006) or distributional generalization (Bolker et al., 2009), as well as dissimilarity-based approaches or spectral decomposition approaches that reduce the number of variables in a statistical analysis (e.g., Legendre and Anderson, 1999; Anderson, 2001; McArdle and Anderson, 2001; Zapala and Schork, 2006; Chiu and Chao, 2014; Warton et al., 2012; Warton and Hudson, 2004) have been frequently applied to omics data. Those methods are being complemented by new statistical approaches allowing to represent and assess configurations of associations (e.g., graph-based approaches; Lima-Mendez et al., 2015) and methods accounting for the compositionality of omics data (e.g., Kurtz et al., 2015; Fernandes et al., 2014; Lovell et al., 2010).

Recently, weighted gene correlation network analysis (WGCNA) has been used to detect significant associations between planktonic species and their genes with carbon export in the ocean at global scale, within the Tara Oceans framework (Guidi et al., 2016). The development of appropriate bioinformatic and statistical tools, coupled to meta-omics, herein allowed to unveil the strong correlation of unexpected taxa (Radiolaria, alveolate parasites, *Synechococcus* and their phages) in oligotrophic regions of the ocean. More importantly, the tool enabled to identify species and gene markers with strong predictive power (using PLS regression) since a significant portion of carbon export variability could be explained. The above mentioned result should prompt additional investigations using similar approaches to investigate the relationships between community structure and diversity with nutrient concentrations, such as the Redfield (1958) ratio.

5. Part III. Semantic approaches in support of conceptual modelling

Before the construction of a model or while organizing data for analysis, the entities of interest – e.g., ecosystem processes, organisms, populations, communities – and their interrelations should be clearly defined. To this end, conceptual models are produced as a guiding precursor to numerical modelling endeavors; however, concepts are often hard to articulate in a robust and reproducible manner, and a loose approach at this early stage can easily contaminate model definition and exchangeability. Models which contend with biotic and abiotic plankton interactions are particularly at risk of such semantic contamination, due to the sheer breadth of interdisciplinarity concepts called upon. In this section, we will discuss the use of semantic methodologies to improve

the specification of conceptual models and expose them in a machine-readable form to pave the way for coherent and automated data and model sharing (Fig. 3).

As we have described above, contemporary plankton modelers must bring to bear knowledge across disciplines such as oceanography, remote sensing, ecology, omics, and microbiology to improve model outputs. Unfortunately, ambiguity lurks within the terminologies and conceptualizations native to these domains, which inhibits coherent integration across research groups. For example, and as discussed above in the context of Follows et al. (2007), a “species” variable in microbial diversity assessment can pertain to very different real-world entities dependent on the methods chosen to generate them (e.g., Koeppl and Wu, 2014; Thompson, 2015; Varghese et al., 2015; and Rosselló-Móra and Amann, 2015). Additionally, the very nature of planktonic “interactions” is often merely a statement of co-occurrence, heavily influenced by how occurrence information has been analyzed and which experts have been called upon to interpret the data. Only sometimes, such as in the case of symbiotic associations in Lima-Mendez et al. (2015; see above), can interactions be validated by direct observation. Further, variables of key interest to plankton dynamics such as “sea ice extent” in the Arctic (e.g., Metfies et al., 2016) may contain markedly different information dependent on measurement methods such as satellite telemetry, aerial and ship-based surveys, and chemical proxies (see, e.g., Vinnikov et al., 1999; Nicol et al., 2000; Wolff et al., 2006). Even conceptualizations of geographically scaled entities such as “biomes”, “ecoregions”, and “provinces” are far from homogeneous (e.g., Gibbons et al., 2013; Dunn et al., 2013; Ulrich and Gotelli, 2010), especially in the aquatic realm (e.g., Spalding et al., 2007; Oliver and Irwin, 2008; Reygondeau et al., 2013). This semantic ambiguity may easily become toxic to both numerical and statistical modelling efforts by introducing a source of variation which is very difficult to trace and correct for. Great manual effort is required to remedy semantic confounding retroactively (e.g. ten Hoopen et al., 2016), which is likely to become even more difficult as the volume and variety of data relevant to plankton modelling increases alongside the need to reliably connect models to monitoring frameworks through reporting standards such as the Essential Ocean Variables (EOVs) (e.g., Lindstrom et al., 2012; Constable et al., 2016). Below, we describe methods and resources that can proactively reduce, or at least control, semantic ambiguity in ecology and that show promise in transparently aligning conceptual models with the real-world entities they aim to describe. This coherence will greatly support the modelling community by addressing the pressing challenges in climate and biodiversity science (Hardisty et al., 2013).

Capturing semantics in human- and machine-usable form is widely accomplished through the development of information artifacts called *ontologies* (for an introduction, see Arp et al., 2015). Generally, ontologies represent entities (objects, regions, processes, etc.) as “classes”, similar to nodes in a network, and the relationships between them as logical expressions (e.g. “subclass of”, “part of”, and “derived from”). Well-constructed ontologies can help reduce the ambiguity associated with variables by explicitly representing the web of meaning behind them using standardized logical expressions. Of course, ontologies cannot solve the issues of limited knowledge or disagreement in a given discipline; however, they are able to explicitly define and represent measured or simulated entities, alongside the methods used to capture information about them, which underpin any given variable. In this mode, they are capable of facilitating the capture and integration of expert knowledge across domains, allowing modelers enhanced insight into how their model components relate to material entities and processes which may influence their behaviour.

Among the various efforts to bring ontology into ecology (e.g., Madin et al., 2008) the Open Biomedical and Biological Ontologies (OBO) Foundry and Library (Smith et al., 2007) is integrating ecological ontologies into its federated set of semantic resources. The OBO community initially addressed entities from the biomedical domain, most notably through the Gene Ontology (GO; The Gene Ontology Consortium,

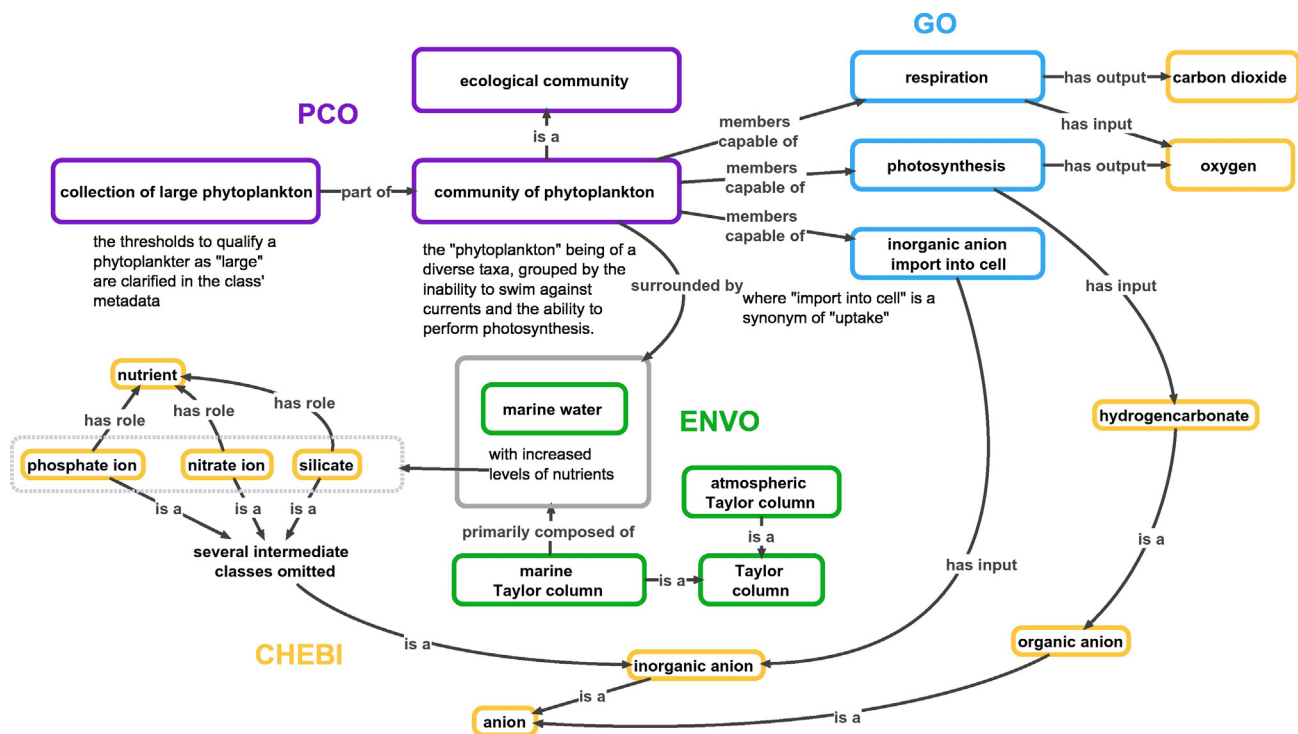


Fig. 3. A simplified scheme of how selected components in the BFM model can be embedded within existing ontologies to reduce ambiguity around key concepts. While much more links are possible, we show a simplified representation of how the Population and Community Ontology, the Gene Ontology, the Chemical Entities of Biological Interest Ontology, and the Environment Ontology can contribute to model clarification (see Section 3 in the main text for more detail). Each node represents a class and each edge represents a logical relation.

2000); however, new OBO ontologies are now representing domains of ecological relevance (for recent review see Thessen et al., 2015). For example, the Environment Ontology (ENVO; Buttigieg et al., 2013, 2016) provides semantic support for representing ecosystems, habitats, environmental materials, environmental processes, and – recently – environmental variables. In our present case, this ontology would allow the specification of entities (geographic features, oceanographic processes, etc.) which are relevant to the hydrodynamics affecting plankton dispersion. ENVO links environmental material entities through the processes they participate in using logical relations, which numerical modelers may enhance with rate estimates or other such qualifiers. Naturally, different groupings of the plankton will have varied interactions with these environmental entities; therefore, resources such as the Population and Community Ontology (PCO; Walls et al., 2014a, 2014b) are needed to link conceptual models with structured, reproducible representations of organism groupings. To illustrate, consider some of the biotic components shown in Fig. 2a: groupings of organisms into categories such as “Large phytoplankton” or “Aerobic bacteria” can easily vary between studies using different classification approaches or thresholds. Creating well-defined classes in the PCO (as is the case with any well-designed ontology) will allow different groups of researchers to declare their definitional thresholds and rationale in a more discoverable and transparent manner, aiding reproducibility and comparability across studies and promoting the FAIR principles (Wilkinson et al., 2016). The representations in the PCO can be readily extended to functional or trait-based groupings (i.e. populations or communities of organisms which share traits or ecological functions) and even greater precision is possible should those phenotypes, themselves, be represented ontologically (see below).

Above, we explore how modelers can more clearly map their conceptualizations of plankton responses to environmental influences using two relevant ontologies. However, if needed, many more ontologies are available to prevent ambiguity creeping into modelling efforts. Remaining in the ecological domain, the Biocollections Ontology (BCO; Walls et al., 2014b) offers semantics to support conceptual modelling

of collecting and sampling processes as well as observational processes. Guidelines on applying these ontologies to OSD records has been described by Walls et al. (2014b) and scope exists to extend their application to similar omics initiatives targeting other ecosystems relevant to plankton research. The semantic representations in the ecologically-themed ontologies discussed above can be combined with existing OBO ontologies to provide fine-grained resources to better define conceptual models (for demonstration of one such combination, see Rocca-Serra et al., 2015). GO is a natural fit to any omically enabled plankton modelling, providing classes which describe the biological processes (e.g. photosynthesis, uptake) of relevance in plankton research (as illustrated in Fig. 3). Further, GO, as with a number of the ontologies mentioned here, can provide additional analytical capacities (e.g. Mi et al., 2013). In addition, ontologies such as the Chemical Entities of Biological Interest (CHEBI; Degtyarenko et al., 2008), the Ontology for Biomedical Investigations (OBI; Brinkman et al., 2010), and the Information Artifact Ontology (IAO) can provide reusable components for modelers to avail of. CHEBI provides a developed representation of the relations between over 50,000 chemical entities which may be used, in combination with ENVO's environmental materials, to express variables such as the concentration and dispersion of beta-lactam antibiotics in seawater. Using CHEBI, the ambiguity within operational groupings such as “semi-labile DOM” (Fig. 3) can be reduced by declaring which compounds or compound classes they include. Planned processes, such as the measuring of variables or the processing of omics data which can introduce their own biases, can be defined with OBI while the plans which specify them can be represented and explicitly identified using IAO. IAO can also be used to declare what data is about (for foundational thinking on “aboutness”, see Smith and Ceusters, 2015) and what kind of digital information is being handled (e.g. a ‘measurement datum’, a ‘setting datum’, or a ‘Cartesian spatial co-ordinate datum’), thereby preventing unintended data confounding. Ontology-assisted conceptual modelling may be further extended to address anatomical parts or organismal products such as fish guts or animal waste (of considerable importance to the microbial ecology and

chemistry of coastal regions, especially near river mouths or heavily populated areas) through the use of anatomy ontologies such as the multi-species Uber Anatomy Ontology (UBERON; Mungall et al., 2012). Such technologies have the potential to marry data standardization, integration, and conceptual modelling at scale in the age of information. Importantly, these standards are flexible and modelers are free to declare new relations between, e.g., environmental processes and different plankton groups and associate them with the corresponding mathematical expressions from their numerical model. Should these prove to be robust, they can be “pushed” back to the main ontologies used as a form of rapid and actionable knowledge dissemination.

New ontologies with further relevance to plankton modelling in the face of global change are emerging, extending the resources available to the modelling community. For example, the Ontology of Microbial Phenotypes (OMP; Chibucos et al., 2014) should be explored in attempts to link microbial phenotypic and taxonomic data to environmental modelling. Further, the global drive for sustainable development (United Nations, 2015) is being represented in collaboration with the United Nations Environment Programme (UNEP) by the Sustainable Development Goals Interface Ontology (SDGIO, <https://github.com/SDG-InterfaceOntology/sdgio>). This will allow research outcomes from projects such as MicroB3 (<https://microb3.eu/>) and AtlantOS (<http://atlantos-h2020.eu/>) to be placed within a policy-aligned frame of reference in a machine-actionable manner and linked to SDGs such as Goal 13: “Take urgent action to combat climate change and its impacts” and Goal 14: “Conserve and sustainably use the oceans, seas and marine resources for sustainable development”. Such an ability to tie conceptual models, data, and model predictions to policy targets in a standardized yet flexible manner is highly desirable in bridging advanced research with decision-making processes.

To realize the potential in linking semantics to modelling, future work should focus on collaboration between ontologists and modelers of all kinds in order to generate project-specific ontologies. These artifacts will both draw from and shape the general-use, domain ontologies described above in order to better expose modeled knowledge in plankton ecology. With such semantic clarity, powerful information resources resembling the Global Biotic Interactions platform (Poelen et al., 2014) and TraitBank (Parr et al., 2014) can be used to discover and refine data and supply both statistical and numerical modelling, as well as their outputs. Such efforts aim to empower analyses such as those undertaken by Henschel et al. (2015), which used the logical structure of ontologies for sophisticated analyses of microbiome diversity data. While there is much to be improved upon, such work demonstrates that the skilled use of ontologies can greatly enhance the encoding, transfer, application and refinement of knowledge about phytoplankton ecology in aid of modelling and data analysis.

6. Conclusions

This review was built around the question: are we ready to integrate omics data, in present biogeochemical and ecological models of plankton?

It is undeniable that the astonishing amount of molecular data generated by meta-omics approaches may theoretically be of the greatest importance to interrogate the dynamics regulating ocean planktonic systems. This body of information will inevitably grow alongside its complexity and variety. However, the use of this data to improve numerical models is still in its infancy. This delay, together with the difficulties in evaluating the reliability of statistical models, are major obstacles in transforming that information into improved understanding of plankton ecology and its responses to pressures such as global warming or pollution-linked impacts. The forecasting of such responses to proposed management policies will also suffer without further improvements in linking information artifacts to semantically clarified representations of policies and planned interventions. We have described a spectrum of research outcomes which are converging in

order to reduce the gap between the multifaceted potential of contemporary giga-sciences and the ability of scientists to synchronize together their data within a more clearly defined and universally accessible theoretical framework. Herein, we have performed an initial synthesis of both present-day methodologies as well as future potentialities which may orient research efforts to develop a new generation of models to predict the evolution of the marine ecosystem. In doing so, we have also illustrated some of the achievements in recent years that highlight the importance of integrating omics data in present and future modelling initiatives. Therefore we may conclude that we are not yet ready but are on the way.

Acknowledgements

KFS, LC, PLB, DI, and CB were funded by the EU project MicroB3. KFS was partially funded by the SZN Open University PhD Program. KFS, LC, MRA, DDA and DI were partially funded by the Flagship Project RITMARE – The Italian Research for the Sea – funded by the Italian Ministry of Education, University, and Research within the National Research Program 2011–2013 and by the 'Progetto Premiale' MIUR NEMO. PLB also acknowledges support from the ERC Advanced Investigator grant ABYSS 294757 to Antje Boetius. CB additionally acknowledges funding from the ERC Advanced Award “Diatomite” and from the Louis D Foundation, as well as the French Government “Investissements d’Avenir” programmes MEMO LIFE (ANR- 10-LABX-54), PSL* Research University (ANR-11-IDEX-0001-02), and OCEANOMICS (ANR-11-BTBR-0008), as well as the Gordon and Betty Moore Foundation. MBS acknowledges funding from a Gordon and Betty Moore Foundation Investigator Award (#3790, to MBS). MBS additionally acknowledges funding from Tara Oceans NSF grant with “NSF OCE#1536989 to MBS”.

References

- Aerts, R., Honnay, O., 2011. Forest restoration, biodiversity and ecosystem functioning. *BMC Ecol.* 11:29. <http://dx.doi.org/10.1186/1472-6785-11-29>.
- Ainsworth, C., Samhouri, J., Busch, D., Cheung, W.W., Dunne, J., Okey, T.A., 2011. Potential impacts of climate change on northeast pacific marine foodwebs and fisheries. *ICES J. Mar. Sci.* 68 (6), 1217–1229.
- Alexander, H., Rouco, M., Haley, S., Wilson, S., Karl, D., Dyhrman, S., 2015. Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc. Natl. Acad. Sci.* E5972–E5979 <http://dx.doi.org/10.1073/pnas.1518165112>.
- Allen, J., Polimene, L., 2011. Linking physiology to ecology: towards a new generation of plankton models. *J. Plankton Res.* 33 (7), 989–997.
- Anderson, D.M., 1998. Physiology and bloom dynamics of toxic *Alexandrium* species, with emphasis on life cycle transitions. *Physiological Ecology of Harmful Algal Blooms*. Springer-Verlag, Berlin, Heidelberg, pp. 19–48.
- Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26 (1), 32–46.
- Anderson, T.R., 2005. Plankton functional type modelling: running before we can walk? *J. Plankton Res.* 27 (3) 1–1. 1, 2.
- Anderson, T.R., Pondaven, P., 2003. Non-redfield carbon and nitrogen cycling in the Sargasso Sea: pelagic imbalances and export flux. *Deep-Sea Res. I Oceanogr. Res. Pap.* 50 (5), 573–591.
- Anning, T., MacIntyre, H.L., Pratt, S.M., Sammes, P.J., Gibb, S., Geider, R.J., 2000. Photoacclimation in the marine diatom *Skeletonema costatum*. *Limnol. Oceanogr.* 45, 1807–1817.
- Armstrong, R.A., 1999. An optimization-based model of iron-light-ammonium colimitation of nitrate uptake and phytoplankton growth. *Limnol. Oceanogr.* 44, 1436–1446.
- Armstrong, R.A., 2006. Optimality-based modelling of nitrogen allocation and photoacclimation in photosynthesis. *Deep-Sea Res. II Top. Stud. Oceanogr.* 53 (5), 513–553.
- Arp, R., Smith, B., Spear, A.D., 2015. *Building Ontologies With Basic Formal Ontology*. The MIT Press, Cambridge, MA, USA.
- Arrigo, K.R., Robinson, D.H., Worthen, D.L., Dunbar, R.B., Di Tullio, G.R., VanWoert, M., Lizotte, M.P., 1999. Phytoplankton community structure and the drawdown of nutrients and CO₂ in the Southern Ocean. *Science* 283 (5400), 365–367.
- Assmy, P., Smetacek, V., Montresor, M., Klaas, C., Henjes, J., Strass, V.H., Arrieta, J.M., Bathmann, U., Berg, G.M., Breitbarth, E., Cisewski, B., Friedrichs, L., Fuchs, N., Herndl, G.J., Jansen, S., Kragefsky, S., Latasa, M., Peeken, I., Rottgers, R., Scharek, R., Schuller, S.E., Steigenberger, S., Webb, A., Wolf-Gladrow, D., 2013. Thick-shelled, grazer-protected diatoms decouple ocean carbon and silicon cycles in the iron-limited Antarctic Circumpolar Current. *Proc. Natl. Acad. Sci.* 110:20633–20638. <http://dx.doi.org/10.1073/pnas.1309345110>.
- Baas Becking, L.G.M., 1934. In: Van Stockum, W.P., Zoon, N.V. (Eds.), *Geobiologie of inleiding tot de milieukunde*. Den Haag [The Netherlands].

- Bachvaroff, T.R., Kim, S., Guillou, L., Delwiche, C.F., Coats, D.W., 2012. Molecular diversity of the syndinean genus *euduboscquella* based on single-cell pcr analysis. *Appl. Environ. Microbiol.* 78 (2), 334–345.
- Bálint, M., Bahram, M., Eren, A.M., Faust, K., Fuhrman, J.A., Lindahl, B., O'Hara, R.B., Öpik, M., Sogin, M.L., Unterseher, M., Tedersoo, L., 2016. Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiol. Rev.* 40:686–700. <http://dx.doi.org/10.1093/femsre/fuw017>.
- Banase, K., 1982. Mass-scaled rates of respiration and intrinsic growth in very small invertebrates. *Mar. Ecol. Prog. Ser.* 9 (3), 281–297.
- Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B., Ferrer, E.A., 2011. Has the Earth's sixth mass extinction already arrived? *Nature* 471:51–57. <http://dx.doi.org/10.1038/nature09678>.
- Barton, A.D., Irwin, A.J., Finkel, Z.V., Stock, C.A., 2016. Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proc. Natl. Acad. Sci.* 113:2964–2969. <http://dx.doi.org/10.1073/pnas.1519080113>.
- Barton, A.D., Pershing, A.J., Litchman, E., Record, N.R., Edwards, K.F., Finkel, Z.V., Kjørboe, T., Ward, B.A., 2013. The biogeography of marine plankton traits. *Ecol. Lett.* 16:522–534. <http://dx.doi.org/10.1111/ele.12063>.
- Becks, L., Ellner, S.P., Jones, L.E., Hairston, N.G., 2012. The functional genomics of an eco-evolutionary feedback loop: linking gene expression, trait evolution, and community dynamics. *Ecol. Lett.* 15 (5), 492–501.
- Behl, S., Donval, A., Stibor, H., 2011. The relative importance of species diversity and functional group diversity on carbon uptake in phytoplankton communities. *Limnol. Oceanogr.* 56:683–694. <http://dx.doi.org/10.4319/lo.2011.56.2.0683>.
- Bergkvist, J., Thor, P., Jakobsen, H.H., Wangberg, S.A., Selander, E., 2012. Grazer-induced chain length plasticity reduces grazing risk in a marine diatom. *Limnol. Oceanogr.* 57:318–324. <http://dx.doi.org/10.4319/lo.2012.57.1.0318>.
- Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., Kiko, R., Not, F., 2016. In situ imaging reveals the biomass of giant protists in the global ocean. *Nature* 532:504–507. <http://dx.doi.org/10.1038/nature17652>.
- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E., 2005. Defining operational taxonomic units using DNA barcode data. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 360:1935–1943. <http://dx.doi.org/10.1098/rstb.2005.1725>.
- Blüthgen, N., Fründ, J., Vázquez, D.P., Menzel, F., 2008. What do interaction network metrics tell us about specialization and biological traits. *Ecology* 89 (12), 3387–3399.
- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24 (3), 127–135.
- Breitbart, M., 2012. Marine viruses: truth or dare. *Annu. Rev. Mar. Sci.* 4:425–448. <http://dx.doi.org/10.1146/annurev-marine-120709-142805>.
- Brinkman, R.R., Courtot, M., Derom, D., Fostel, J.M., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., Rutenberg, A., Sansone, S.-A., Soldatova, L.N., Stoeckert Jr., C.J., Turner, J.A., Zheng, J., the OBI consortium, 2010. Modelling biomedical experimental processes with OBI. *J. Biomed. Semantics* 1 (Suppl. 1):S7. <http://dx.doi.org/10.1186/2041-1480-1-S1-S7>.
- Bruggeman, J., Kooijman, S.A.L.M., 2007. A biodiversity-inspired approach to aquatic ecosystem modelling. *Limnol. Oceanogr.* 52, 1533–1544.
- Brum, J.R., Sullivan, M.B., 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* 13 (3), 147–159.
- Brum, J.R., Hurwitz, B.L., Schofield, O., Ducklow, H.W., Sullivan, M.B., 2015a. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J.* 1–13.
- Brum, J.R., Ignacio-espinoza, J.C., Roux, S., Doulier, G., Acinas, S.G., Alberti, A., Chaffron, S., 2015b. Ocean viral communities. *Science* 348 (1261498):1–11. <http://dx.doi.org/10.1126/science.1261498>.
- Brum, J.R., Ignacio-Espinoza, J.C., Kim, E.H., Trubl, G., Jones, R.M., Roux, S., VerBruggen, N.C., Rich, V.I., Sullivan, M.B., 2016. Illuminating structural proteins in viral “dark matter” with metaproteomics. *Proc. Natl. Acad. Sci.* 113 (9), 2436–2441.
- Buttigieg, P.L., Ramette, A., 2014. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiol. Ecol.* 90:543–550. <http://dx.doi.org/10.1111/1574-6941.12437>.
- Buttigieg, P.L., Ramette, A., 2015. Biogeographic patterns of bacterial microdiversity in Arctic deep-sea sediments (HAUSGARTEN, Fram Strait). *Front. Microbiol.* 5, 1–12.
- Buttigieg, P.L., Morrison, N., Smith, B., Mungall, C.J., Lewis, S.E., 2013. The environment ontology: contextualising biological and biomedical entities. *J. Biomed Semantics* 4 (1), 43.
- Buttigieg, P.L., Pafilis, E., Lewis, S.E., Schildhauer, M.P., Walls, R.L., Mungall, C.J., 2016. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperability. *J. Biomed. Semantics* 7:57. <http://dx.doi.org/10.1186/s13326-016-0097-6>.
- Cao, K.A.L., Costello, M.E., Lakis, V.A., Bartolo, F., Chua, X.Y., Brazeilles, R., Rondeau, P., 2016. MixMC: a multivariate statistical framework to gain insight into microbial communities. *PLoS One* 11:1–30. <http://dx.doi.org/10.1371/journal.pone.0160169>.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.L., Huttley, G.A., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7 (5), 335–336.
- Cardinale, B.J., Matulich, K.L., Hooper, D.U., Byrnes, J.E., Duffy, E., Gamfeldt, L., Balvanera, P., O'Connor, M.I., Gonzalez, A., 2011. The functional role of producer diversity in ecosystems. *Am. J. Bot.* 98 (3), 572–592.
- Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naem, S., 2012. Biodiversity loss and its impact on humanity. *Nature* 486 (7401), 59–67 2.
- Carmona, C.P., de Bello, F., Mason, N.W., Lepš, J., 2016. Traits without borders: integrating functional diversity across scales. *Trends Ecol. Evol.* 31, 382–394.
- Caron, D., Countway, P., 2009. Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat. Microb. Ecol.* 57 (3), 227–238.
- Caron, D.A., Connell, P.E., Schaffner, R.A., Schnetzer, A., Fuhrman, J.A., Countway, P.D., Kim, D.Y., 2017. Planktonic food web structure at a coastal time-series site: I. Partitioning of microbial abundances and carbon biomass. *Deep-Sea Res. I Oceanogr. Res. Pap.* <http://dx.doi.org/10.1016/j.dsr.2016.12.013>.
- Chaffron, S., Rehner, H., Pernthaler, J., von Mering, C., 2010. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* 20 (7), 947–959.
- Chambouvet, A., Alves-de Souza, C., Cuffe, V., Marie, D., Karpov, S., Guillou, L., 2011. Interplay between the parasite *amoebophrya* sp. (alveolata) and the cyst formation of the red tide dinoflagellate *Scrippsiella trochoidea*. *Protist* 162 (4), 637–649.
- Chesson, P., Donahue, M.J., Melbourne, B., Sears, A.L., 2005. Scale transition theory for understanding mechanisms in metacommunities. In: Holyoak, M., Leibold, M.A., Holt, R.D. (Eds.), *Metacommunities: Spatial Dynamics and Ecological Communities*. University of Chicago Press, Chicago, IL, pp. 279–306.
- Chibucos, M.C., Zweifel, A.E., Herrera, J.C., Meza, W., Eslamfam, S., Uetz, P., Siegle, D.A., Hu, J.C., Giglio, M.G., 2014. An ontology for microbial phenotypes. *BMC Microbiol.* 14:294. <http://dx.doi.org/10.1186/s12866-014-0294-3>.
- Chiu, C.-H., Chao, A., 2014. Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers. *PLoS One* 9, e100014.
- Clark, C.M., Flynn, D.F.B., Butterfield, B.J., Reich, P.B., 2012. Testing the link between functional diversity and ecosystem functioning in a Minnesota grassland experiment. *PLoS One* 7, e25821.
- Coelho, S.M., Simon, N., Ahmed, S., Cock, J.M., Partensky, F., 2013. Ecological and evolutionary genomics of marine photosynthetic organisms. *Mol. Ecol.* 22 (3), 867–907.
- Constable, A.J., Costa, D.P., Schofield, O., Newman, L., Urban, E.R., Fulton, E.A., Melbourne-Thomas, J., Ballerini, T., Boyd, P.W., Brandt, A., de la Mare, W.K., Edwards, M., Eléaume, M., Emmerson, L., Fennel, K., Fielding, S., Griffiths, H., Gutt, J., Hindell, M.A., Hofmann, E.E., Jennings, S., La, H.S., McCurdy, A., Mitchell, B.G., Moltmann, T., Muelbert, M., Murphy, E., Press, A.J., Raymond, B., Reid, K., Reiss, C., Rice, J., Salter, I., Smith, D.C., Song, S., Southwell, C., Swadling, K.M., Van de Putte, A., Willis, Z., 2016. Developing priority variables (ecosystem Essential Ocean Variables - eEOVs) for observing dynamics and change in Southern Ocean ecosystems. *J. Mar. Syst.* 161:26–41. <http://dx.doi.org/10.1016/j.jmarsys.2016.05.003>.
- Cottenie, K., Meester, L.D., 2003. Comment to Oksanen (2001): reconciling Oksanen (2001) and Hurlbert (1984). *Oikos* 100 (2):394–396. <http://dx.doi.org/10.1034/j.1600-0706.2003.11953.x>.
- Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., Arnaud-Haond, S., 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS One* 10:1–26. <http://dx.doi.org/10.1371/journal.pone.0117562>.
- Crise, A., Kaberi, H., Ruiz, J., Zetsepina, A., Arashkevich, E., Giani, M., Karageorgis, A.P., Prieto, L., Pantazi, M., Gonzalez-Fernandez, D., Ribera d'Alcalá, M., Tornero, V., Vassilopoulou, V., Durrieu de Madron, X., Guieu, C., Puig, P., Zenetos, A., Andral, B., Angel, D., Altkuhov, D., Ayata, S.D., Aktan, Y., Balcioglu, E., Benedetti, F., Bouchoucha, M., Buia, M.-C., Cadiou, J.-F., Canals, M., Chakroun, M., Christou, E., Christidis, M.G., Civitarese, G., Coatu, V., Corsini-Foka, M., Cozzi, S., Deidun, A., Dell'Aquila, A., Dogrammatzi, A., Dumitracu, C., Edelist, D., Ettahiri, O., Fonda-Umani, S., Gana, S., Galgani, F., Gasparini, S., Giannakourou, A., Gomoiu, M.-T., Gubanova, A., Gucu, A.-C., Gurses, O., Hanke, G., Hatzianestis, I., Herut, B., Hone, R., Huertas, E., Irissou, J.-O., Isinibilir, M., Jimenez, J.A., Kalogiour, S., Kapiris, K., Karamfilov, V., Kavadas, S., Keskis, C., Kideys, A.E., Kocak, M., Kondylatos, G., Kontogiannis, C., Kosyan, R., Koubbi, P., Kusplilic, G., La Ferla, R., Langone, L., Laroche, S., Lazar, L., Lefkaditou, E., Lemeschko, I.E., Machias, A., Malej, A., Mazzocchi, M.-G., Medinets, V., Mihalopoulos, N., Miserochci, S., Moncheva, S., Mukhanov, V., Oaie, G., Oros, A., Ozturk, A.A., Ozturk, B., Panayotova, M., Prospathopoulos, A., Radu, G., Raykov, V., Reglero, P., Reygondeau, G., Rougeron, N., Salihoglu, B., Sanchez-Vidal, A., Sannino, G., Santinelli, C., Scerriou, D., Shapiro, G., Simbora, N., Shiganova, T., Sprovieri, M., Stefanova, K., Streftaris, N., Tirelli, V., Tom, M., Topalogu, B., Topu, N.E., Tsagarakis, K., Tsangaris, C., Tserpes, G., Tugrul, S., Uysal, Z., Vasile, D., Violaki, K., Xu, J., Yuksek, A., Papathanassiou, E., 2015. A MSFD complementary approach for the assessment of pressures, knowledge and data gaps in Southern European Seas: the PERSEUS experience. *Mar. Pollut. Bull.* 95, 25–39.
- Cristescu, M.E., 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* 29, 566–571.
- Daines, S.J., Clark, J.R., Lenton, T.M., 2014. Multiple environmental controls on phytoplankton growth strategies determine adaptive responses of the N:P ratio. *Ecol. Lett.* 17 (4), 414–425.
- Dakos, V., Benincà, E., van Nes, E.H., Philippart, C.J.M., Scheffer, M., Huisman, J., 2009. Inter-annual variability in species composition explained as seasonally entrained chaos. *Proc. Biol. Sci.* 276:2871–2880. <http://dx.doi.org/10.1098/rspb.2009.0584>.
- D'Alelio, D., Ribera d'Alcalá, M., Dubroca, L., Sarno, D., Zingone, A., Montresor, M., 2010. The time for sex: a biennial life cycle in a marine planktonic diatom. *Limnol. Oceanogr.* 55 (1), 106–114.
- D'Alelio, D., Mazzocchi, M.G., Montresor, M., Sarno, D., Zingone, A., Margiotta, F., Saggiomo, V., Di Capua, I., Franz, G., Ribera, M., 2015. The green – blue swing: plasticity of plankton food-webs in response to coastal oceanographic dynamics. *Mar. Ecol.* 36:1155–1170. <http://dx.doi.org/10.1111/maec.12211>.
- D'Alelio, D., Libralato, S., Wyatt, T., Ribera d'Alcalá, M., 2016. Ecological-network models link diversity, structure and function in the plankton food-web. *Sci. Rep.* 6:21806. <http://dx.doi.org/10.1038/srep21806>.
- Darzi, Y., Falony, G., Vieira-Silva, S., Raes, J., 2015. Towards biome-specific analysis of meta-omics data. *ISME J.* 10, 1025–1028.
- David, W., Archibald, J.M., 2016. Evolution: plumbing the depths of diplomid diversity. *Curr. Biol.* 26 (24), 1290–1292.

- Davies, N., Field, D., Amaral-Zettler, L., Clark, M., Deck, J., Drummond, A., Faith, D., Geller, J., Gilbert, J., Glockner, F., Hirsch, P., Leong, J.-A., Meyer, C., Obst, M., Planes, S., Scholin, C., Vogler, A., Gates, R., Toonen, R., Berteaux-Lecellier, V., Barbier, M., Barker, K., Bertillon, S., Bickel, M., Bietz, M., Bobe, J., Bodrossy, L., Borja, A., Coddington, J., Fuhrman, J., 2014. The founding charter of the genomic observatories network. *Gigascience* 3:2. <http://dx.doi.org/10.1186/2047-217X-3-2>.
- Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F., Taberlet, P., 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Lett.* 10: 1789–1793. <http://dx.doi.org/10.1098/rsbl.2014.0562>.
- Decelle, J., Probert, I., Bittner, L., Desdèves, Y., Colin, S., de Vargas, C., Galí, M., Simó, R., Not, F., 2012. An original mode of symbiosis in open ocean plankton. *Proc. Natl. Acad. Sci.* 109 (44), 18000–18005.
- Degtyarenko, K., De Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M., Ashburner, M., 2008. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* 36:344–350. <http://dx.doi.org/10.1093/nar/gkm791>.
- Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., Sullivan, M.B., 2014. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* 513 (7517), 242–245.
- Droop, M.R., 1974. The nutrient status of algal cells in continuous culture. *J. Mar. Biol. Assoc. U. K.* 54, 825–855.
- Duffy, J.E., Stachowicz, J.J., 2006. Why biodiversity is important to oceanography: potential roles of genetic, species, and trophic diversity in pelagic ecosystem processes. *Mar. Ecol. Prog. Ser.* 311:179–189. <http://dx.doi.org/10.3354/meps311179>.
- Duhaime, M.B., Deng, L., Poulos, B.T., Sullivan, M.B., 2012. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* 14 (9), 2526–2537.
- Dunlop, E.S., Heino, M., Dieckmann, U., 2009. Eco-genetic modelling of contemporary life-history evolution. *Ecol. Appl.* 19 (7), 1815–1834.
- Dunn, R.R., Fierer, N., Henley, J.B., Leff, J.W., Menninger, H.L., 2013. Home life: factors structuring the bacterial diversity found within and between homes. *PLoS One* 8 (5), e64133.
- Edlund, A., Yang, Y., Yooseph, S., Hall, A.P., Nguyen, D.D., Dorrestein, P.C., Nelson, K.E., He, X., Lux, R., Shi, W., McLean, J.S., 2015. Meta-omics uncover temporal regulation of pathways across oral microbiome meta during in vitro sugar metabolism. *ISME J.* 9:2605–2619. <http://dx.doi.org/10.1038/ismej.2015.72>.
- Edwards, R.A., Rohwer, F., 2005. Viral metagenomics. *Nat. Rev. Microbiol.* 3 (6), 504–510.
- Edwards, K.F., Thomas, M.K., Klausmeier, C.A., Litchman, E., 2012. Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. *Limnol. Oceanogr.* 57 (2), 554–566 2.
- Edwards, K.F., Litchman, E., Klausmeier, C.A., 2013. Functional traits explain phytoplankton community structure and seasonal dynamics in a marine ecosystem. *Ecol. Lett.* 16 (1), 56–63.
- Elliott, J.A., Irish, A.E., Reynolds, C.S., 2001. The effects of vertical mixing on a phytoplankton community: a modelling approach to the intermediate disturbance hypothesis. *Freshw. Biol.* 46, 1291–1297.
- Ellner, S.P., Geber, M.A., Hairston, N.G., 2011. Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics. *Ecol. Lett.* 14 (6), 603–614.
- Esposito, S., Botte, V., Iudicone, D., Ribera d'Alcalà, M.R., 2009. Numerical analysis of cumulative impact of phytoplankton photoresponses to light variation on carbon assimilation. *J. Theor. Biol.* 261, 361–371.
- Estep, K.W., Davis, P.G., Keller, M.D., Sieburth, J.M., 1986. How important are oceanic algal nanoflagellates in bacterivory? *Limnol. Oceanogr.* 31 (3), 646–650 1.
- Evans, G.T., Parslow, J.S., 1985. A model of annual plankton cycles. *Biol. Oceanogr.* 3, 327–347.
- Evans, T.G., Chan, F., Menge, B.A., Hofmann, G.E., 2013. Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. *Mol. Ecol.* 22, 1609–1625.
- Falcitatore, A., Bowler, C., 2002. Revealing the molecular secrets of marine diatoms. *Annu. Rev. Plant Biol.* 53 (1), 109–130.
- Falkowski, P.G., Barber, R.T., Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* 281 (5374), 200–206.
- Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10 (8), 538–550.
- Fernandes, A.D., Reid, J.N.S., Macklaim, J.M., Mcmurrough, T.A., Edgell, D.R., Gloor, G.B., 2014. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2, 15.
- Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281 (5374), 237–240.
- Finkel, Z.V., 2001. Light absorption and size scaling of light-limited metabolism in marine diatoms. *Limnol. Oceanogr.* 46 (1), 86–94.
- Finkel, Z.V., Irwin, A.J., 2000. Modelling size-dependent photosynthesis: light absorption and the allometric rule. *J. Theor. Biol.* 204 (3), 361–369.
- Flatt, T., Heyland, A. (Eds.), 2011. *Mechanisms of Life History Evolution. The Genetics and Physiology of Life History Traits and Trade-Offs*. OUP, Oxford, p. 506.
- Flegontova, O., Flegontov, P., Malviya, S., Audic, S., Wincker, P., de Vargas, C., Bowler, C., Lukes, J., Horák, A., 2016. Extreme diversity of diplomonad eukaryotes in the ocean. *Curr. Biol.* 26:3060–3065. <http://dx.doi.org/10.1016/j.cub.2016.09.031>.
- Follows, M.J., Dutkiewicz, S., 2011. Modelling diverse communities of marine microbes. *Annu. Rev. Mar. Sci.* 3:427–451. <http://dx.doi.org/10.1146/annurev-marine-120709-142848>.
- Follows, M.J., Dutkiewicz, S., Grant, S., Chisholm, S.W., 2007. Emergent biogeography of microbial communities in a model ocean. *Science* 315 (5820), 1843–1846.
- Franzosa, E.A., 2015. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat. Rev. Microbiol.* 13, 360–372.
- Fuhrman, J.A., 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* 399 (6736), 541–548.
- Fuhrman, J.A., Cram, J.A., Needham, D.M., 2015. Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* 13:133–146. <http://dx.doi.org/10.1038/nrmicro3417u>.
- Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R., Gilbert, J.A., 2013. Evidence for a persistent microbial seed bank throughout the global ocean. *Proc. Natl. Acad. Sci.* 110 (12), 4651–4655.
- Gifford, S.M., Sharma, S., Moran, M.A., 2014. Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community. *Front. Microbiol.* 5:1–12. <http://dx.doi.org/10.3389/fmicb.2014.00185>.
- Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., Weynberg, K., Huse, S., Hughes, M., Joint, I., Somerfield, P.J., Muhling, M., 2010. The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic' study of seasonal and diel temporal variation. *PLoS One* 5. <http://dx.doi.org/10.1371/journal.pone.0015545>.
- Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlami, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J., Coelho, L.P., Espinoza, J.C.I., Malviya, S., Sunagawa, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Poulain, J., Searson, S., Coordinators, T.O.C., Stemmann, L., Not, F., Hingamp, P., Speich, S., Follows, M., Karp-Boss, L., Boss, E., Ogata, H., Pesant, S., Weissenbach, J., Wincker, P., Acinas, S.G., Bork, P., de Vargas, C., Iudicone, D., Sullivan, M.B., Raes, J., Karsenti, E., Bowler, C., Gorsky, G., 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532, 465–470.
- Guillou, L., Viprey, M., Chambouvet, A., Welsh, R., Kirkham, A., Massana, R., Scanlan, D.J., Worden, A., 2008. Widespread occurrence and genetic diversity of marine parasitoids belonging to syndiniales (alveolata). *Environ. Microbiol.* 10 (12), 3349–3365.
- Haeckel, E.H.P.A., 1866. *Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanische Begründet durch die von Charles Darwin reformirte Descendenz-Theorie. Volume I: Allgemeine Anatomie der Organismen. 32 + 574 pages; volume II: Allgemeine Entwicklungsgeschichte der Organismen. 140 + 462 pages.* Georg Reimer, Berlin, Germany.
- Haggerty, J.M., Dinsdale, E.A., 2016. Distinct biogeographical patterns of marine bacterial taxonomy and functional genes. *Glob. Ecol. Biogeogr.* 26:177–190. <http://dx.doi.org/10.1111/geb.12528>.
- Hairston, N.G.J., Hairston, N.G.S., 1993. Cause-effect relationships in energy flow, trophic structure, and interspecific interactions. *Am. Nat.* 142, 379–411.
- Hardisty, A., Roberts, D., Biodiversity Informatics Community, 2013. A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecol.* 13 <http://dx.doi.org/10.1186/1472-6785-13-16>.
- Hartmann, M., Grob, C., Tarran, G.A., Martin, A.P., Burkill, P.H., Scanlan, D.J., Zubkov, M.V., 2012. Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proc. Natl. Acad. Sci.* 109 (15), 5756–5760.
- Hellweger, F.L., Clegg, R.J., Clark, J.R., Plugge, C.M., Kreft, J.-U., 2016. Advancing microbial sciences by individual-based modelling. *Nat. Rev. Microbiol.* 14, 461–471.
- Hendry, A.P., Kinnison, M.T., 1999. Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53, 1637–1653.
- Henschel, A., Anwar, M.Z., Manohar, V., 2015. Comprehensive meta-analysis of ontology annotated 16S rRNA profiles identifies Beta diversity clusters of environmental bacterial communities. *PLoS Comput. Biol.* 11 (10), e1004468.
- Hilton, J.A., Foster, R.A., Tripp, H.J., Carter, B.J., Zehr, J.P., Villareal, T.A., 2013. Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. *Nat. Commun.* 4, 1767.
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., 2015. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* <http://dx.doi.org/10.1093/nar/pkv1248>.
- Hurlbert, S.H., 2004. On misinterpretations of pseudoreplication and related matters: a reply to Oksanen. *Oikos* 104 (3), 591–597.
- Hurwitz, B.L., Westveld, A.H., Brum, J.R., Sullivan, M.B., 2014. Modelling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proc. Natl. Acad. Sci. U. S. A.* 111:10714–10719. <http://dx.doi.org/10.1073/pnas.1319778111>.
- Hutchinson, G.E., 1957. Cold Spring Harbor Symposium on Quantitative Biology. Concluding Remarks. 22 pp. 415–427.
- Hutchinson, G.E., 1961. The paradox of the plankton. *Am. Nat.* 95 (882), 137–145.
- Ichihashi, Y., Aguilar-Martínez, J.A., Farhi, M., Chitwood, D.H., Kumar, R., Millon, L.V., Peng, J., Maloof, J.N., Sinha, N.R., 2014. Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proc. Natl. Acad. Sci. U. S. A.* 111:E2616–E2621. <http://dx.doi.org/10.1073/pnas.1402835111>.
- Ioannidis, J.P.A., 2005. Why most published research findings are false. *PLoS Med.* 2 (8), e124.
- Ioannidis, J.P.A., 2014. How to make more published research true. *PLoS Med.* 11. <http://dx.doi.org/10.1371/journal.pmed.1001747>.
- Irwin, A.J., Finkel, Z.V., Schofield, O.M., Falkowski, P.G., 2006. Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. *J. Plankton Res.* 28 (5), 459–471.
- Irwin, A.J., Finkel, Z.V., Müller-Karger, F.E., Troccoli Ghinaglia, L., 2015. Phytoplankton adapt to changing ocean environments. *Proc. Natl. Acad. Sci. U. S. A.* 1–5. <http://dx.doi.org/10.1073/pnas.1414752112>.

- Jeong, H.J., Du Yoo, Y., Kim, J.S., Seong, K.A., Kang, N.S., Kim, T.H., 2010. Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. *Ocean Sci. J.* 45 (2), 65–91.
- Jordán, F., 2009. Keystone species and food webs. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 364:1733–1741. <http://dx.doi.org/10.1098/rstb.2008.0335>.
- Kawamiya, M., Kishi, M.J., Sugino, N., 2000. An ecosystem model for the North Pacific embedded in a general circulation model: part I: model description and characteristics of spatial distributions of biological variables. *J. Mar. Syst.* 25 (2), 129–157.
- Kim, M., Rai, N., Zorraqino, V., Tagkopoulos, I., 2016. Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*. *Nat. Commun.* 7, 13090.
- Koeppl, A.F., Wu, M., 2014. Species matter: the role of competition in the assembly of congeneric bacteria. *ISME J.* 8 (3), 531–540.
- Kooijman, S.A., 2001. Quantitative aspects of metabolic organization: a discussion of concepts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356 (1407), 331–349.
- Kooijman, S.A., 2010. *Dynamic Energy Budget Theory for Metabolic Organisation*. third ed. Cambridge University Press, Cambridge, p. 508.
- Koostira, W., Gersonde, R., Medlin, L.K., Mann, D.G., Falkowski, P., Knoll, A., 2007. The origin and evolution of the diatoms: their adaptation to a planktonic existence. *Evolution of Primary Producers in the Sea*, pp. 207–249.
- Kopf, A., Bick, M., Kottmann, R., Schnetzer, J., Kostadinov, I., Lehmann, K., Fernandez-Guerra, A., Jeanthon, C., Rahav, E., Ullrich, M., Wichels, A., 2015. The ocean sampling day consortium. *GigaScience* 4 (1), 1.
- Kostka, J.E., Teske, A.P., Joye, S.B., Head, I.M., 2014. The metabolic pathways and environmental controls of hydrocarbon biodegradation in marine ecosystems. *Front. Microbiol.* 5, 471.
- Krause, S., Le Roux, X., Niklaus, P.A., van Bodegom, P.M., Lennon, J.T., Bertilsson, S., Grossart, H.P., Philippot, L., Bodelier, P.L.E., 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front. Microbiol.* 5: 1–10. <http://dx.doi.org/10.3389/fmicb.2014.00251>.
- Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bonneau, R.A., 2015. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* 11 (5), e1004226.
- Labonté, J.M., Swan, B.K., Poulos, B., Luo, H., Koren, S., Hallam, S.J., Sullivan, M.B., Woyke, T., Wommack, K.E., Stepanauskas, R., 2015. Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *ISME J.* 9 (11), 2386–2399.
- Laliberté, E., Tylianakis, J.M., 2010. Deforestation homogenizes tropical parasitoid–host networks. *Ecology* 91, 1740–1747.
- Laws, E.A., 1975. The importance of respiration losses in controlling the size distribution of marine phytoplankton. *Ecology* 56 (2), 419–426.
- Le Quere, C., Harrison, S.P., Colin Prentice, I., Buitenhuis, E.T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K.E., Legendre, L., Manizza, M., Platt, T., Rivkin, R.B., Sathyendranath, S., Uitz, J., Watson, A.J., Wolf-Gladrow, D., 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Glob. Chang. Biol.* 11:2016–2040. <http://dx.doi.org/10.1111/j.1365-2486.2005.01004.x>.
- Legendre, P., Anderson, M.J., 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* 69 (1), 1–24.
- Legendre, P., Legendre, L., 2012. *Numerical Ecology*. third ed. Elsevier, Amsterdam.
- Leray, M., Knowlton, N., 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proc. Natl. Acad. Sci.* 112 (7), 2076–2081.
- Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignaciopinos, J.C., Roux, S., Vincent, F., Bittner, L., 2015. Determinants of community structure in the global plankton interactome. *Science* 348:1262073. <http://dx.doi.org/10.1126/science.1262073>.
- Lindstrom, E., Gunn, J., Fischer, A., McCurdy, A., Glover, L.K., 2012. A Framework for Ocean Observing: Prepared by the Task Team for an Integrated Framework for Sustained Ocean Observing (IFSOO). (Report IOC/INF-1284 Rev.). UNESCO, Paris, p. 28.
- Litchman, E., Klausmeier, C.A., 2008. Trait-based community ecology of phytoplankton. *Annu. Rev. Ecol. Syst.* 39 (1), 615–639.
- Litchman, E., Klausmeier, C.A., Schofield, O.M., Falkowski, P.G., 2007. The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. *Ecol. Lett.* 10 (12), 1170–1181.
- Lovell, D., Warren, M., Taylor, J., Zwart, A., Helliwell, C., 2010. Caution! Compositions! Can Constraints on Omics Data Lead Analyses astray? Report EP10994.
- Madin, J.S., Bowers, S., Schildhauer, M.P., Jones, M.B., 2008. Advancing ecological research with ontologies. *Trends Ecol. Evol.* 23 (3):159–168. <http://dx.doi.org/10.1016/j.tree.2007.11.007>.
- Madin, J.S., Hoogenboom, M.O., Connolly, S.R., Darling, E.S., Falster, D.S., Huang, D., Keith, S.A., Mizerek, T., Pandolfi, J.M., Putnam, H.M., Baird, A.H., 2016. A trait-based approach to advance coral reef science. *Trends Ecol. Evol.* 31 (6), 419–428.
- Májeková, M., Paal, T., Plovan, N.S., Bryndová, M., Kasari, L., Norberg, A., Weiss, M., Bishop, T.R., Luke, S.H., Sam, K., Le Bagousse-Pinguet, Y., 2016. Evaluating functional diversity: missing trait data and the importance of species abundance structure and data transformation. *PLoS One* 11 (2), e0149270.
- Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Ludicone, D., de Vargas, C., Bittner, L., Zingone, A., Bowler, C., 2016. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci.* 113: E1516–E1525. <http://dx.doi.org/10.1073/pnas.1509523113>.
- Marchetti, A., Schrueth, D.M., Durkin, C.A., Parker, M.S., Kodner, R.B., Berthiaume, C.T., Morales, R., Allen, A.E., Armbrust, E.V., 2012. PNAS plus: comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl. Acad. Sci.* 109:E317–E325. <http://dx.doi.org/10.1073/pnas.1118408109>.
- Margalef, R., 1974. *Ecología*. Omega, Barcelona, p. 951.
- Margalef, R., 1978. Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanol. Acta* 1 (4), 493–509.
- McArdle, B., Anderson, M.J., 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82 (1), 290–297.
- McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8 (4), e61217. <http://dx.doi.org/10.1371/journal.pone.0061217>.
- Metfies, K., von Appen, W.J., Kilias, E., Nicolaus, A., Nöthig, E.M., 2016. Biogeography and photosynthetic biomass of Arctic marine pico-eukaryotes during summer of the record sea ice minimum 2012. *PLoS One* 11 (2), e0148512.
- Mi, H., Muruganujan, A., Casagrande, J.T., Thomas, P.D., 2013. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* 8 (8), 1551–1566.
- Mizuno, C.M., Rodriguez-Valera, F., Kimes, N.E., Ghai, R., 2013. Expanding the marine virosphere using metagenomics. *PLoS Genet.* 9 (12), e1003987.
- Mock, T., Kirkham, A., 2012. What can we learn from genomics approaches in marine ecology? From sequences to eco-systems biology! *Mar. Ecol. Prog. Ser.* 33 (2), 131–148.
- Moore, L.R., Post, A.F., Rocap, G., Chisholm, S.W., 2002. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* 47 (4), 989–996.
- Morrissey, C.A., Mineau, P., Devries, J.H., Sanchez-Bayo, F., Liess, M., Cavallaro, M.C., Liber, K., 2015. Neonicotinoid contamination of global surface waters and associated risk to aquatic invertebrates: a review. *Environ. Int.* 74, 291–303.
- Motti, C., 2012. Environmental marine metabolomics: from whole organism system biology to ecosystem management. *J. Marine Sci. Res. Dev.* 2, 3.
- Mouchet, M.A., Villéger, S., Mason, N.W.H., Moullot, D., 2010. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Funct. Ecol.* 24, 867–876.
- Mungall, C.J., Tornai, C., Gkoutos, G.V., Lewis, S.E., Haendel, M.A., 2012. Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* 13 (1), R5.
- Naeem, S., Kawabata, Z., Loreau, M., 1998. Transcending boundaries in biodiversity research. *Trends Ecol. Evol.* 13:134–135 Personal Ed. 13. 10.1016/S0169-5347(97)01316-5.
- Needham, D.M., Chow, C.E., Cram, J.A., Sachdeva, R., Parada, A., Fuhrman, J.A., 2013. Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. *ISME J.* 7, 1274–1285.
- Nicol, S., Pauly, T., Bindoff, N.L., Wright, S., Thiele, D., Hosie, G.W., Strutton, P.G., Woehler, E., 2000. Ocean circulation off east Antarctica affects ecosystem structure and sea-ice extent. *Nature* 406:504–507. <http://dx.doi.org/10.1038/35020053>.
- Nuzzo, R., 2014. Statistical errors. *Nature* 506:150–152. <http://dx.doi.org/10.1136/bmj.1.6053.66>.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. 27 pp. 29–34.
- Oksanen, L., 2001. Logic of experiments in ecology: is pseudoreplication a pseudoissue? *Oikos* 94 (1), 27–38.
- Oliver, M.J., Irwin, A.J., 2008. Objective global ocean biogeographic provinces. *Geophys. Res. Lett.* 35 (15):1–6. <http://dx.doi.org/10.1029/2008GL034238>.
- Palmer, J., Totterdell, I., 2001. Production and export in a global ocean ecosystem model. *Deep-Sea Res. I Oceanogr. Res. Pap.* 48 (5), 1169–1198.
- Parr, C.S., Wilson, N., Schulz, K., Leary, P., Hammock, J., Rice, J., Corrigan Jr., R.J., 2014. TraitBank: practical semantics for organism attribute data. *Semant Web Interoperability Usability Appl.* 650–1860.
- Penczykowski, R.M., Forde, S.E., Duffy, M.A., 2011. Rapid evolution as a possible constraint on emerging infectious diseases. *Freshw. Biol.* 56 (4), 689–704.
- Pocock, M.J.O., Evans, D.M., Memmott, J., 2012. The robustness and restoration of a network of ecological networks. *Science* 335, 973–977.
- Poelen, J.H., Simons, J.D., Mungall, C.J., 2014. Global biotic interactions: an open infrastructure to share and analyze species-interaction datasets. *Eco. Inform.* 24, 148–159.
- Polimene, L., Brunet, C., Butenschön, M., Martínez-Vicente, V., Widdicombe, C., Torres, R., Allen, J.L., 2013. Modelling a light-driven phytoplankton succession. *J. Plankton Res.* 36, 214–229.
- Prosser, J.I., 2010. Replicate or lie. *Environ. Microbiol.* 12 (7), 1806–1810.
- Proulx, S.R., Promislow, D.E.L., Phillips, P.C., 2005. Network thinking in ecology and evolution. *Trends Ecol. Evol.* 20:345–353. <http://dx.doi.org/10.1016/j.tree.2005.04.004>.
- Ptácnik, R., Solimini, A.G., Andersen, T., Tamminen, T., Brettum, P., Lepistö, L., Willén, E., Rekolainen, S., 2008. Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *Proc. Natl. Acad. Sci. U. S. A.* 105 (13), 5134–5138.
- Redfield, A.C., 1958. The biological control of chemical factors in the environment. *Am. Sci. pages* 230A–221. 2.
- Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., Maury, O., 2013. Dynamic biogeochemical provinces in the global ocean. *Glob. Biogeochem. Cycles* 27 (4):1046–1058. <http://dx.doi.org/10.1002/gbc.20089>.
- Reynolds, C.S., 1984. *The Ecology of Freshwater Phytoplankton*. Cambridge University Press, p. 1.
- Richardson, A.J., Schoeman, D.S., 2004. Climate impact on plankton ecosystems in the Northeast Atlantic. *Science* 305 (5690), 1609–1612.
- Rocca-Serra, P., Walls, R., Parnell, J., Gallery, R., Zheng, J., Sansone, S.-A., Gonzalez-Beltran, A., 2015. Modelling a Microbial Community and Biodiversity assay With OBO Foundry Ontologies: The Interoperability Gains of a Modular Approach Database 2015(0), bau132.
- Rosselló-Móra, R., Amann, R., 2015. Past and future species definitions for Bacteria and Archaea. *Syst. Appl. Microbiol.* 38 (4), 209–216.
- Roux, S., Hallam, S.J., Woyke, T., Sullivan, M.B., 2015. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *life* 4, e08490.
- Sarà, G., Palmeri, V., Rinaldi, A., Montalto, V., Helmuth, B., 2013. Predicting biological invasions in marine habitats through eco-physiological mechanistic models: a case study with the bivalve *Brachidontes pharaonis*. *Divers. Distrib.* 19, 1235–1247.

- Sarà, G., Rinaldi, A., Montalto, V., 2014. Thinking beyond organism energy use: a trait-based bioenergetic mechanistic approach for predictions of life history traits in marine organisms. *Mar. Ecol. Prog. Ser.* 463, 506–515.
- Saraiva, S., van der Meer, J., Kooijman, S.A., Witbaard, R., Philippart, C.J.M., Hippler, D., Parker, R., 2012. Validation of a dynamic energy budget (DEB) model for the blue mussel *Mytilus edulis*. *Mar. Ecol. Prog. Ser.* 463, 141–158.
- Sardans, J., Penuelas, J., Rivas-Ubach, A., 2011. Ecological metabolomics: overview of current developments and future challenges. *Chemoecology* 21 (4), 191–225.
- Satinsky, B.M., Crump, B.C., Smith, C.B., Sharma, S., Zielinski, B.L., Doherty, M., Meng, J., Sun, S., Medeiros, P.M., Paul, J.H., Coles, V.J., Yager, P.L., Moran, M.A., 2014a. Microspatial gene expression patterns in the Amazon River Plume. *Proc. Natl. Acad. Sci. U. S. A.* 111:11085–11090. <http://dx.doi.org/10.1073/pnas.1402782111>.
- Satinsky, B.M., Zielinski, B.L., Doherty, M., Smith, C.B., Sharma, S., Paul, J.H., Crump, B.C., Moran, M.A., 2014b. The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome* 2:17. <http://dx.doi.org/10.1186/2049-2618-2-17>.
- Sauterey, B., Ward, B.A., Follows, M.J., Bowler, C., Claessen, D., 2015. When everything is not everywhere but species evolve: an alternative method to model adaptive properties of marine ecosystems. *J. Plankton Res.* 37, 28–47.
- Schartau, M., Oschlies, A., 2003. Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: part I - method and parameter estimates. *J. Mar. Res.* 61 (6), 765–793.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541.
- Selander, E., Jakobsen, H.H., Lombard, F., Kiorboe, T., 2011. Grazer cues induce stealth behavior in marine dinoflagellates. *Proc. Natl. Acad. Sci.* 108, 4030–4034.
- Seymour, J.R., Doblin, M.A., Jeffries, T.C., Brown, M.V., Newton, K., Ralph, P.J., Baird, M., Mitchell, J.G., 2012. Contrasting microbial assemblages in adjacent water masses associated with the East Australian Current. *Environ. Microbiol. Rep.* 4, 548–555.
- Shi, Y., Tyson, G.W., Eppley, J.M., DeLong, E.F., 2011. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J.* 5, 999–1013.
- Sieracki, M.E., Verity, P.G., Stoecker, D.K., 1993. Plankton community response to sequential silicate and nitrate depletion during the 1989 North Atlantic spring bloom. *Deep-Sea Res. II Top. Stud. Oceanogr.* 40 (1), 213–225.
- Smetacek, V., 2012. Making sense of ocean biota: how evolution and biodiversity of land organisms differ from that of the plankton. *J. Biosci.* 37:589–607. <http://dx.doi.org/10.1007/s12038-012-9240-4>.
- Smith, B., Ceusters, W., 2015. Aboutness: towards foundations for the information artifact ontology. *Proceedings of the 6th International Conference on Biomedical Ontology*, July 26–30. Lisbon, Portugal, pp. 47–51.
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S., 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25:1251–1255. <http://dx.doi.org/10.1038/nbt1346>.
- Solonenko, S.A., Sullivan, M.B., 2013. In: Delong, E.F. (Ed.), *Methods in Enzymology: Microbial Community "Omics": Metagenomics, Metatranscriptomics, and Metaproteomics*. Elsevier.
- Soltwedel, T., Schauer, U., Boebel, O., Nothig, E.M., Bracher, A., Metfies, K., ... Klages, M., 2013. FRAM - Frontiers in Arctic marine monitoring visions for permanent observations in a gateway to the Arctic Ocean. *OCEANS 2013 MTS/IEEE: The Challenges of the Northern Dimension*. Bergen. <http://dx.doi.org/10.1109/OCEANS-Bergen.2013.6608008>.
- Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Louri, S.A., Martin, K.D., Mcmanus, E., Molnar, J., Recchia, C.A., Robertson, J., 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* 57:773. <http://dx.doi.org/10.1641/B570707>.
- Steele, J.H., 1974. *The Structure of Marine Ecosystems*. Harvard University Press, Cambridge.
- Sterner, R.W., Elser, J.J., 2002. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*. Princeton Univ. Press.
- Stoecker, D.K., Gustafson, D.E., 2003. Cell-surface proteolytic activity of photosynthetic dinoflagellates. *Aquat. Microb. Ecol.* 30 (2), 175–183.
- Sul, W.J., Oliver, T.A., Ducklow, H.W., Amaral-Zettler, L.A., Sogin, M.L., 2013. Marine bacteria exhibit a bipolar distribution. *Proc. Natl. Acad. Sci. U. S. A.* 110, 2342–2347.
- Sullivan, M.B., 2015. Viromes, not gene markers for studying dsDNA virus communities. *J. Virol.* 89:2459–2461. <http://dx.doi.org/10.1128/JVI.03289-14>.
- Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., Cornejo-Castillo, F.M., Costea, P.I., Cruaud, C., D'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J.M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B.T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemann, L., Sullivan, M.B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S.G., Bork, P., 2015. Structure and function of the global ocean microbiome. *Science* 348:1261359. <http://dx.doi.org/10.1126/science.1261359>.
- Sutherland, W.J., Freckleton, R.P., Godfray, H.C.J., Beissinger, S.R., Benton, T., Cameron, D.D., Carmel, Y., Coomes, D.A., Coulson, T., Emmerson, M.C., Hails, R.S., Hays, G.C., Hodgson, D.J., Hutchings, M.J., Johnson, D., Jones, J.P.G., Keeling, M.J., Kokko, H., Kunin, W.E., Lambin, X., Lewis, O.T., Malhi, Y., Mieszkowska, N., Milner-Gulland, E.J., Norris, K., Phillimore, A.B., Purves, D.W., Reid, J.M., Reuman, D.C., Thompson, K., Travis, J.M.J., Turnbull, L.A., Wardle, D.A., Wiegand, T., 2013. Identification of 100 fundamental ecological questions. *J. Ecol.* 101:58–67. <http://dx.doi.org/10.1111/1365-2745.12025>.
- Suttle, C.A., 2007. Marine viruses – major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812.
- Suweis, S., Simini, F., Banavar, J.R., Maritan, A., 2013. Emergence of structural and dynamical properties of ecological mutualistic networks. *Nature* 500, 449–452.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045–2050.
- Tang, E.P., 1995. The allometry of algal growth rates. *J. Plankton Res.* 17 (6), 1325–1335.
- Tëmkin, I., Eldredge, N., 2015. *Networks and hierarchies: approaching complexity in evolutionary theory*. Macroevo. Springer, pp. 183–226.
- ten Hoopen, P., Amid, C., Buttigieg, P.L., Pafilis, E., Bravakos, P., Cerdeño-Tarraga, A.M., Gibson, R., Kahlke, T., Legaki, A., Murthy, K.N., Papastefanou, G., Pereira, E., Rossello, M., Toribio, A.L., Cochrane, G., 2016. Value, but high costs in post-deposition data Curation. *Database* 2016:1–10. <http://dx.doi.org/10.1093/database/bav126>.
- Tett, P., Gowen, R.J., Painting, S.J., Elliott, M., Forster, R., Mills, D.K., Bresnan, E., Capuzzo, E., Fernandes, T., Foden, J., Geider, R., Gilpin, L., Huxham, M., McQuatters-Gollop, A., Malcolm, S., Saux Picart, S., Platt, T., Racault, M.-F., Sathyendranath, S., van der Molen, J., Wilkinson, M., Geider, R.J., 2013. Framework for understanding marine ecosystem health. *Mar. Ecol. Prog. Ser.* 494, 1–27.
- The Gene Ontology Consortium, 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25:25–29. <http://dx.doi.org/10.1038/75556>.
- Thessen, A.E., Bunker, D.E., Buttigieg, P.L., Cooper, L.D., Dahdul, W.M., Domisch, S., Franz, N.M., Jaiswal, P., Lawrence-Dill, C.J., Midford, P.E., Mungall, C.J., Ramírez, M.J., Specht, C.D., Vogt, L., Vos, R.A., Walls, R.L., White, J.W., Zhang, G., Deans, A.R., Huala, E., Lewis, S.E., Mabee, P.M., 2015. Emerging semantics to link phenotype and environment. *PeerJ* 3, e1470. <http://dx.doi.org/10.7717/peerj.1470>.
- Thingstad, T.F., Cuevas, L.A., 2010. Nutrient pathways through the microbial food web: principles and predictability discussed, based on five different experiments. *Aquat. Microb. Ecol.* 61:249–260. <http://dx.doi.org/10.3354/ame01452>.
- Thompson, F.L., 2015. Microbial taxonomy in the post-genomic era: rebuilding from scratch? *Arch. Microbiol.* 197 (3), 359–370.
- Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladysz, S., Kitching, R.L., Martinez, N.D., Rantala, H., Romanuk, T.N., Stouffer, D.B., Tylianakis, J.M., 2012. Food webs: reconciling the structure and function of biodiversity. *Trends Ecol. Evol.* 27:689–697. <http://dx.doi.org/10.1016/j.tree.2012.08.005>.
- Thuiller, W., Münckmüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffrers, K., Gravel, D., 2013. A road map for integrating eco-evolutionary processes into biodiversity models. *Ecol. Lett.* 16 (s1), 94–105.
- Tilman, D., Kilham, S.S., Kilham, P., 1982. Phytoplankton community ecology: the role of limiting nutrients. *Annu. Rev. Ecol. Syst.* 349–372.
- Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., Siemann, E., 1997. The influence of functional diversity and composition on ecosystem processes. *Science* 277 (5330), 1300–1302.
- Tittel, J., Bissinger, V., Zippel, B., Gaedke, U., Bell, E., Lorke, A., Kamjunke, N., 2003. Mixotrophs combine resource use to outcompete specialists: implications for aquatic food webs. *Proc. Natl. Acad. Sci.* 100 (22), 12776–12781.
- Totterdell, I.J., Armstrong, R., Drange, H., Parslow, J., Powell, T., Taylor, A., 1993. Trophic resolution. Towards a model of ocean biogeochemical processes. *NATO ASI* 1, 10 1.
- Ulrich, W., Gotelli, N.J., 2010. Null model analysis of species associations using abundance data. *Ecology* 91 (11), 3384–3397.
- United Nations, 2015. *Transforming Our World: The 2030 Agenda for Sustainable Development A/RES/70/1*.
- Urich, T., Lanzén, A., Stokke, R., Pedersen, R.B., Bayer, C., Thorseth, I.H., Schleper, C., Steen, I.H., Øvreas, L., 2014. Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated metagenomics. *Environ. Microbiol.* 16:2699–2710. <http://dx.doi.org/10.1111/1462-2920.12283>.
- Valentine, J.W., Jablonski, D., Krug, A.Z., Berke, S.K., 2013. The sampling and estimation of marine paleodiversity patterns: implications of a Pliocene model. *Paleobiology* 39, 1–20.
- Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.M., Peroux, T., Crivelli, A.J., Olivier, A., Acquerberg, M., Le Brun, M., Moller, P.R., Willerslev, E., Dejean, T., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.* 25:929–942. <http://dx.doi.org/10.1111/mec.13428>.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E., 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1–12. <http://dx.doi.org/10.1007/s13398-014-0173-2>.
- Varghese, N.J., Mukherjee, S., Ivanova, N., Konstantinidis, K.T., Mavrommatis, K., Kyrpides, N.C., Pati, A., 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 43:6761–6771. <http://dx.doi.org/10.1093/nar/gkv657>.
- Vasas, V., Jordán, F., 2006. Topological keystone species in ecological interaction networks: considering link quality and non-trophic effects. *Ecol. Model.* 196, 365–378.
- Veldhoen, N., Ikonou, M.G., Helbing, C.C., 2012. Molecular profiling of marine fauna: integration of omics with environmental assessment of the world's oceans. *Ecotoxicol. Environ. Saf.* 76, 23–38.

- Vichi, M., Pinardi, N., Masina, S., 2007. A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory. *J. Mar. Syst.* 64, 89–109.
- Villar, E., Farrant, G.K., Follows, M., Garczarek, L., Speich, S., Audic, S., Bittner, L., Blanke, B., Brum, J.R., Brunet, C., Casotti, R., Chase, A., Dolan, J.R., D'Ortenzio, F., Gattuso, J.-P., Grima, N., Guidi, L., Hill, C.N., Jahn, O., Jamet, J.-L., Le Goff, H., Lepoivre, C., Malviya, S., Pelletier, E., Romagnan, J.-B., Roux, S., Santini, S., Scalco, E., Schwenck, S.M., Tanaka, A., Testor, P., Vannier, T., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S.G., Bork, P., Boss, E., de Vargas, C., Gorsky, G., Ogata, H., Pesant, S., Sullivan, M.B., Sunagawa, S., Wincker, P., Karsenti, E., Bowler, C., Not, F., Hingamp, P., Iudicone, D., 2015. Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* 348:1–12. <http://dx.doi.org/10.1126/science.1261447>.
- Vinnikov, K.Y., Robock, A., Stouffer, R.J., Walsh, J.E., Parkinson, C.L., Cavalieri, D.J., Mitchell, J.F.B., Garrett, D., Zakharov, V.F., 1999. Global warming and northern hemisphere sea ice extent. *Science* 286, 1934–1937.
- Von Dassow, M., Miller, C.J., Davidson, L.A., 2014. Biomechanics and the thermotolerance of development. *PLoS One* 9.
- Walls, R.L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., Bowers, S., Buttigieg, P.L., Davies, N., Endresen, D., Gandolfo, M.A., 2014a. Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies. 9:pp. 1–13. <http://dx.doi.org/10.1371/journal.pone.0089606>.
- Walls, R.L., Guralnick, R., Deck, J., Buntzman, A., Buttigieg, P.L., Davies, N., Denslow, M.W., Gallery, R.E., Parnell, J.J., Osumi-Sutherland, D., Robbins, R.J., Rocca-Serra, P., Wiczorek, J., Zheng, J., 2014b. Meeting report: advancing practical applications of biodiversity ontologies. *Stand. Genomic Sci.* 9:17. <http://dx.doi.org/10.1186/1944-3277-9-17>.
- Ward, B.A., Follows, M.J., 2016. Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proc. Natl. Acad. Sci.* 113 (11), 2958–2963.
- Warton, D.I., Hudson, H.M., 2004. A MANOVA statistic is just as powerful as distance-based statistics for multivariate abundances. *Ecology* 85 (3), 858–874.
- Warton, D.I., Wright, S.T., Wang, Y., 2012. Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* 3 (1), 89–101.
- Weinbauer, M.G., 2004. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28 (2), 127–181.
- Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., Birmingham, A., Cram, J.A., Fuhrman, J.A., Raes, J., Sun, F., Zhou, J., Knight, R., Birmingham, A., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 10, 1669–1681.
- Weitz, J.S., Stock, C.A., Wilhelm, S.W., Bourouiba, L., Coleman, M.L., Buchan, A., Follows, M.J., Fuhrman, J.A., Jover, L.F., Lennon, J.T., Middelboe, M., Sonderegger, D.L., Suttle, C.A., Taylor, B.P., Thingstad, T.F., Wilson, W.H., Wommack, K.E., 2015. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J.* 9:1352–1364. <http://dx.doi.org/10.1038/ismej.2014.220>.
- Wigington, C.H., Sonderegger, D., Brussaard, C.P.D., Buchan, A., Finkel, J.F., Fuhrman, J.A., Wilmes, P., Simmons, S.L., Denef, V.J., Banfield, J.F., 2009. The dynamic genetic repertoire of microbial communities. *FEMS Microbiol. Rev.* 33, 109–132.
- Wilkins, D., Lauro, F.M., Williams, T.W., Demaree, M.Z., Brown, M.V., Hoffman, J.M., Andrews-Pfannkuch, C., McQuaid, J.B., Riddle, M.J., Rintoul, S.R., Cavicchioli, R., 2012. Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. *Environ. Microbiol.* 15, 1318–1333.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. <http://dx.doi.org/10.1038/sdata.2016.18>.
- Wolff, E.W., Fischer, H., Fundel, F., Ruth, U., Twarloh, B., Littot, G.C., Mulvaney, R., Röthlisberger, R., De Angelis, M., Boutron, C.F., Hansson, M., Jonsell, U., Hutterli, M.A., Lambert, F., Kaufmann, P., Stauffer, B., Stocker, T.F., Steffensen, J.P., Bigler, M., Siggaard-Andersen, M.L., Udisti, R., Becagli, S., Castellano, E., Severi, M., Wagenbach, D., Barbante, C., Gabrielli, P., Gaspari, V., 2006. Southern Ocean sea-ice extent, productivity and iron flux over the past eight glacial cycles. *Nature* 440:491–496. <http://dx.doi.org/10.1038/nature06271>.
- Wroblewski, J.S., Sarmiento, J.L., Flierl, G.R., 1988. An ocean basin scale model of plankton dynamics in the North Atlantic: 1. Solutions for the climatological oceanographic conditions in May. *Glob. Biogeochem. Cycles* 2, 199–218.
- Wulff, T., Bauerfeind, E., von Appen, W.J., 2016. Physical and ecological processes at a moving ice edge in the Fram Strait as observed with an AUV. *Deep-Sea Res. I Oceanogr. Res. Pap.* 115, 253–264.
- Yachi, S., Loreau, M., 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl. Acad. Sci.* 96 (4), 1463–1468.
- Yoshida, T., Jones, L.E., Ellner, S.P., Fussmann, G.F., Hairston, N.G., 2003. Rapid evolution drives ecological dynamics in a predator–prey system. *Nature* 424 (6946), 303–306.
- Zapala, M.A., Schork, N.J., 2006. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proc. Natl. Acad. Sci. U. S. A.* 103:19430–19435. <http://dx.doi.org/10.1073/pnas.0609333103>.
- Zepeda Mendoza, M.L., Sicheritz-Pontén, T., Gilbert, M.T.P., 2014. Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. *Brief. Bioinform.* 16:745–758. <http://dx.doi.org/10.1093/bib/bbv001>.
- Zimmermann, J., Abarca, N., Enk, N., Skibbe, O., Kusber, W.H., Jahn, R., 2014. Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. *PLoS One* 9. <http://dx.doi.org/10.1371/journal.pone.0108793>.
- Zou, H., Hastie, T., Tibshirani, R., 2006. Sparse principal component analysis. *J. Comput. Graph. Stat.* 15 (2), 265–286.