

# Uncovering hidden worlds of ocean biodiversity

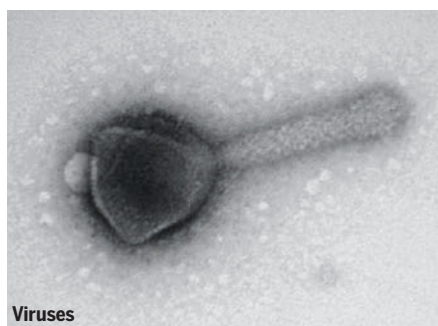
A 3-year expedition yields a treasure trove of data on microorganisms and small animals in the world's oceans

By E. Virginia Armbrust<sup>1</sup>  
and Stephen R. Palumbi<sup>2</sup>

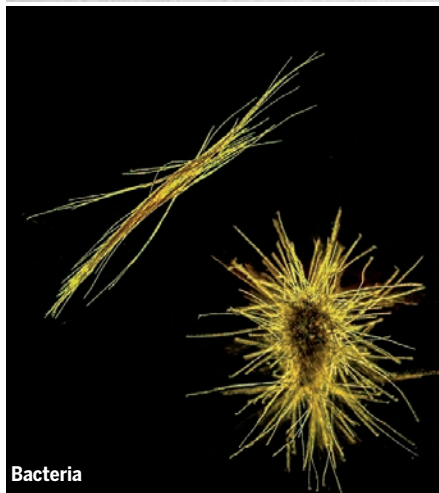
A bewildering swirl of tiny creatures dominates life in the oceans. More numerous than the stars in the universe, these organisms serve as the foundation of all marine food webs, recycling major elements and producing and consuming about half the organic matter generated on Earth each year (1). In this issue, five research articles from the *Tara* Oceans expedition (2–6) provide a vivid, potentially transformative view of the genetic diversity and interconnectivity of these unseen marine communities of viruses, bacteria, archaea, single-celled eukaryotes, and small planktonic animals (see the figure). Together, these studies deliver compelling evidence for extensive networks of previously hidden biological interactions in the sea.

The *Tara* Oceans expedition harkens back to 18th-century sailing voyages that explored uncharted worlds, including Darwin's voyage aboard the *HMS Beagle* and the *Challenger* expedition that heralded the beginning of modern oceanography. The 36-m schooner *Tara* departed Lorient, France, in 2008 and sailed through the Mediterranean Sea and into the Indian, South Atlantic, and Southern oceans (see the map). *Tara* visited coral reefs in the South Pacific Ocean and then sailed through the Panama Canal and back across the North Atlantic Ocean, arriving at her homeport nearly 3 years later. At hundreds of locations along the way, scientists and crew collected thousands of samples from surface waters, from the deep chlorophyll maximum layers where microscopic photosynthetic organisms accumulate, and from deeper waters.

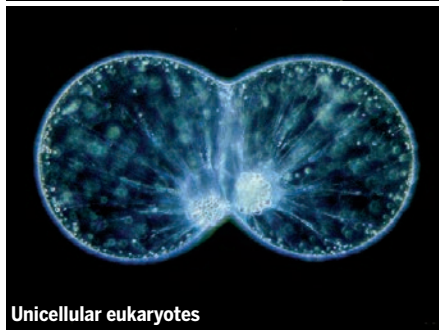
The researchers partitioned the seawater samples into seven size classes, ranging from the smallest viruses to animals less than 2 mm in size. This is where similarities to bygone voyages end: Hundreds of researchers



Viruses



Bacteria



Unicellular eukaryotes



Zooplankton

from around the world used modern DNA sequencing, state-of-the-art microscopy, and computational analyses to examine underlying patterns and drivers of biodiversity.

Most small organisms in the sea cannot be grown and studied in the laboratory and are known only by DNA sequence barcodes—16S ribosomal gene sequences for bacteria and archaea and 18S ribosomal gene sequences for eukaryotes. Even less is known about the genetic makeup of these organisms. Whole-genome sequences are available for the relatively few cultured representatives. Computational approaches (7) and single-cell genomics (8) have generated whole-genome sequences for a select few of the vast majority of marine organisms that remain uncultured. About 10 years ago, the *Sorcerer II* Global Ocean Survey (GOS) first obtained DNA sequences from entire communities of marine microbes, uncovering previously unknown species and genes across the global ocean (9). Building on these results, the *Tara* Oceans expedition has added to the global genetic database a remarkable 7.2 trillion bases of metagenomic DNA sequence from viruses, bacteria, archaea, and those eukaryotes less than ~3 μm in size. In addition, it has yielded millions of 16S and 18S ribosomal bar codes derived from organisms ranging in size from 0.2 μm to 2 mm.

An important outcome of the expedition is the creation by Sunagawa *et al.* (page 873) (2) of an Ocean Microbial Reference Gene Catalog: a collection of more than 40 million nonredundant genes that are blueprints for metabolic function. Given the staggering amount of new sequence information, it is perhaps not surprising that Sunagawa *et al.* found relatively little overlap with the more restricted GOS gene catalog and even less with available reference sequences, emphasizing the likely enormous reservoir of unexplored genetic diversity in these communities. This work also serves as a reminder that the genetic content of domesticated laboratory organisms can differ greatly from that of the abundant taxa found in the wild (10). The data set undoubtedly holds numerous additional whole-genome sequences that await computational reconstruction. Sunagawa

**Ocean diversity.** During the *Tara* Oceans expedition, scientists studied ocean biodiversity in thousands of samples from the world's oceans. From top to bottom, the images show an aquatic virus; filament colonies formed by the cyanobacterium *Trichodesmium* (colony diameter: about 2 mm); the single-cell dinoflagellate *Noctiluca scintillans* (sea sparkle; typical diameter: 0.5 mm) undergoing cell division; and the about 1.5-mm-long zooplankton *Appendicularian Oikopleura dioica*. Five reports in this issue report results from the expedition (2–6).

<sup>1</sup>School of Oceanography, University of Washington, Seattle, WA 98195, USA. <sup>2</sup>Hopkins Marine Station, Stanford University, Stanford, CA 93950, USA.  
E-mail: armbrust@uw.edu; spalumbi@stanford.edu

*et al.* estimate that a few tens of thousands of 16S-based taxa inhabit the upper ocean, similar to earlier predictions from the International Census of Marine Microbes (11). Ocean temperature appears to be the main factor driving these distributions.

Categorizing viruses in the oceans poses a special challenge. Viruses lack a universal molecular identifier, and only a tiny fraction of viruses can be grown in the laboratory. Brum *et al.* (page 874) (3) focused on a subset of DNA-based viruses to generate a “viral pan metagenome.” Based on comparisons with other studies (12), they suggest that the extent of viral genetic diversity in the upper ocean appears well sampled and consists of nearly 1.5 million proteins. As in Sunagawa *et al.*'s study, there was little overlap between their data set and the genetic composition of cultured viral representatives. Although viral community patterns are also influenced by temperature, the biggest driver appears to be related to regional environmental conditions that support seed populations from localized hosts, which are then distributed more broadly via ocean currents.

Despite the impressive computational approaches used in the expedition, apply-

ing metagenomics to larger eukaryotic organisms remains a daunting task, in part because few reference genomes exist. Instead, metabarcoding approaches focus on extensive sampling of target genes, such as the gene encoding cytochrome oxidase c (COI), recently used to identify marine benthic animals (13). In this vein, de Vargas *et al.* (page 874) (4) focused their efforts on generating an extensive database of 18S ribosomal bar codes. They first compiled and generated a comprehensive family tree based on available data and then compared their ~2.3 million bar codes with sequences from different branches of the tree. They estimate that the sunlit regions of the ocean harbor ~150,000 DNA-based taxa of small eukaryotes (less than about 2 mm in size), again providing upper limits for taxonomic diversity and highlighting the difference from reference sequences. The authors find that diversity is greatest within three poorly known groups of unicellular eukaryotes: the Alveolata, Rhizaria, and Excavata. Each consists largely of parasites, phagotrophs (cells that engulf other cells), and symbionts. This observation provides strong evidence that organism interactions drive diversification in marine plankton.

Lima-Mendez *et al.* (page 874) (5) amplify the results of de Vargas *et al.* by providing statistical evidence that parasitic and viral interactions strongly affect population structure. They predict a potential network of interactions between specific organisms that they refer to as an interactome. Their visual confirmation of one such predicted symbiosis between a flatworm and a microalgal suggests that many more novel interactions are embedded in these networks.

Villar *et al.* (page 875) (6) used natural “seas within seas” to study how environmental changes affect marine communities over time and space. The fast-moving Agulhas current hugs the tip of South Africa before slamming into the Antarctic Circumpolar current, a collision that causes most of the Agulhas to turn back on itself to rejoin the Indian Ocean. A small proportion of the current continues westward, however, slipping around Africa into the South Atlantic in the form of giant rotating eddies—the Agulhas rings. These rings, with diameters of hundreds of kilometers, remain physically distinct from surrounding waters and are recognizable in satellite images as they slowly move across the Atlantic toward South America. Villar *et al.* document the

## Exploring hidden ocean worlds, the modern way.

This map shows the route taken by the *Tara* Ocean expedition, which sampled microorganisms and small animals in the world's oceans between 2009 and 2012.

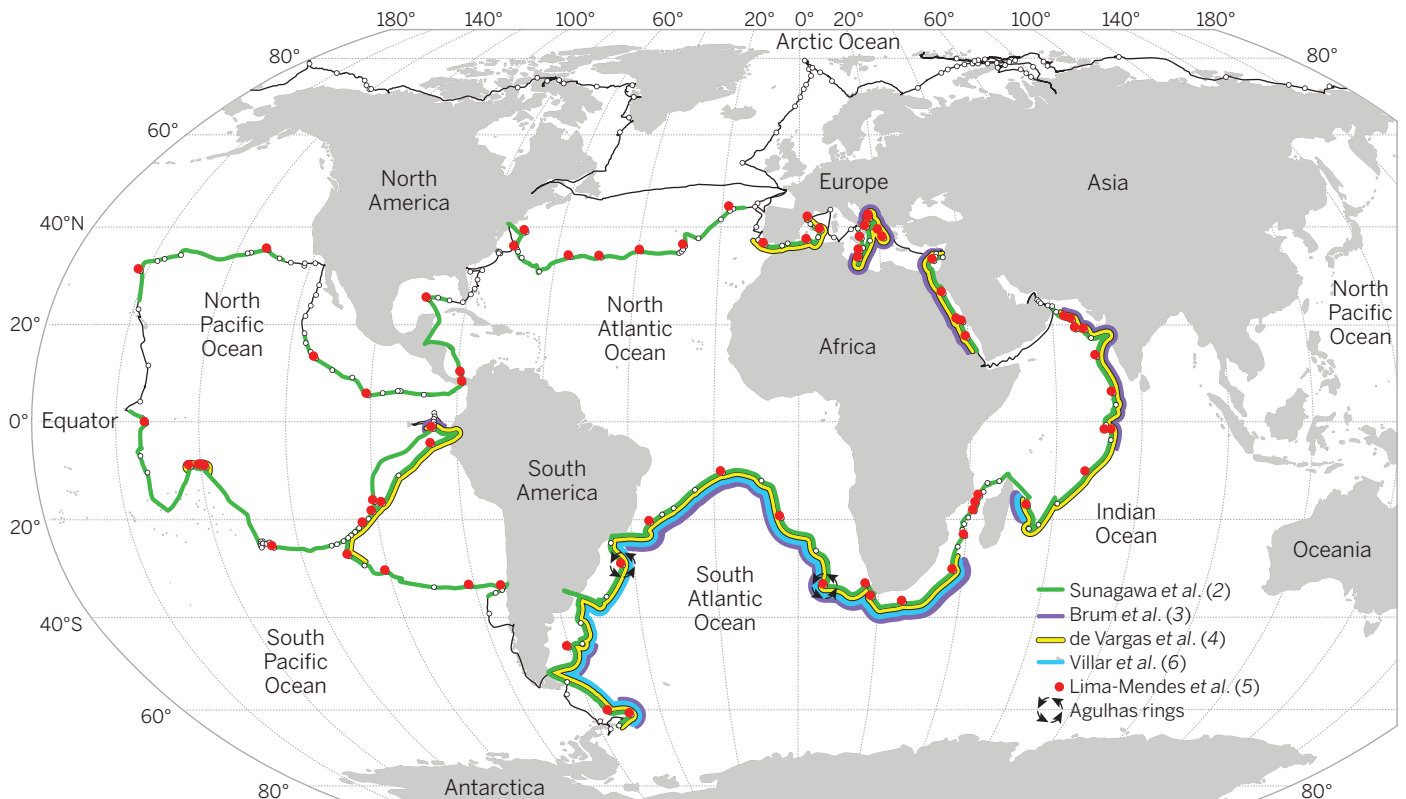


ILLUSTRATION: N. LEBESCOIT AND C. DE VARGAS/IFEP-CNRS ROSCOFF

fate of the communities trapped in these rings. They show that the rings are leaky incubators, changing in oceanographic features and nutrient structure on their trek across the Atlantic. Taxonomic composition changes dramatically along the way, emphasizing the importance of environmental factors. As a result, the rings do not inject plankton from the Indian Ocean into the Atlantic Ocean as previously hypothesized (14).

The studies illustrate the exquisite complexity of marine ecosystems, in which microscopic organisms interact through competition for limiting resources, predator-prey and parasite-host dependencies, and cross-kingdom synergies. These interactions are embedded in a backdrop of fluctuating temperatures, light, and nutrient concentrations. In this world, viscous rather than gravitational forces dominate, allowing heterogeneous hotspots to develop along microscapes (15). Some organisms compete for essential metabolites dissolved in the water, whereas others likely rely on various forms of trading alliances. Yet others simply surround small organic particles, consuming them as they drift to the ocean bottom as “marine snow.” These diverse interactions across a large number of different species raise the question of whether coevolution acts largely between pairs of closely interacting species or on many species interacting within consortia.

The greatest challenge will be to uncover unifying principles behind these interactions. What key currencies are exchanged, how do organisms recognize one another, and when are abiotic conditions more important than species interactions in determining distribution and abundance? How and when should this complexity be incorporated into ecosystem models? When do these interactions affect climate tipping points for ecosystem function? The *Tara* Oceans expedition has generated a treasure trove of data available to anyone willing to dive in and start addressing these questions. ■

#### REFERENCES

1. F. Azam, F. Malfatti, *Nat. Rev. Microbiol.* **5**, 782 (2007).
2. S. Sunagawa *et al.*, *Science* **348**, 1261359 (2015).
3. J. R. Brum *et al.*, *Science* **348**, 1261498 (2015).
4. C. de Vargas *et al.*, *Science* **348**, 1261605 (2015).
5. G. Lima-Mendez *et al.*, *Science* **348**, 1262073 (2015).
6. E. Villar *et al.*, *Science* **348**, 1261447 (2015).
7. V. Iversen *et al.*, *Science* **335**, 587 (2012).
8. H. S. Yoon *et al.*, *Science* **332**, 714 (2011).
9. J. C. Venter *et al.*, *Science* **304**, 66 (2004).
10. C. Rinke *et al.*, *Nature* **499**, 431 (2013).
11. L. Zinger *et al.*, *PLOS ONE* **6**, e24570 (2011).
12. B. L. Hurwitz, M. B. Sullivan, *PLOS ONE* **8**, e57355 (2013).
13. M. Leray, N. Knowlton, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2076 (2015).
14. P. Cermeño, P. G. Falkowski, *Science* **325**, 1539 (2009).
15. R. Stocker, *Science* **338**, 628 (2012).

10.1126/science.aaa7378

## CANCER

# Preprocancer

## Normal skin harbors cancer-causing mutations

By Douglas E. Brash

**T**umors have a life history (1, 2). A tumor consists of a dominant cell clone containing mutations in key cancer genes called “drivers,” together with smaller clones that descended from it but then diverged by accumulating different drivers. Other mutant clones appeared and vanished according to the selection pressures acting upon them or by neutral drift; the resulting diversity confers the means to escape clinical treatment (3). The stepwise accumulation of genetic and epigenetic alterations roughly parallels the clinical progression from normal tissue to precancer, cancer, and metastasis. It is widely assumed that driver mutations occur infrequently in long-lived lineages of cells (2), and that most arise in cancerous tissue that is too small to be clinically detectable. On page 880 of this issue, Martincorena *et al.* (4) overthrow both assumptions and reveal that sun-exposed normal skin is already a polyclonal quilt of driver mutations subjected to selection—a field of preprocancers, as it were—that nevertheless functions as a skin.

The prevailing model for the evolution of cancer is a reiterative process of clonal expansion, genetic diversification, and clonal selection. This multiple-genetic-hit model implies that normal tissue contains a few cells that have driver mutations, but not enough drivers in any one cell to create a cancer or precancer. Indeed, clones of cells with mutations in the tumor suppressor gene *P53* are found in histologically normal human skin and breast tissue; in skin, they constitute a surprising 4% of the epidermis (5, 6).

Martincorena *et al.* used ultradeep sequencing technology to detect other mutations in normal skin. They examined 74 genes in eyelids, a sun-exposed tissue often removed in plastic surgery, and examined the spatial distribution of the mutations by subdividing four lids into 234 minibiopsies. It turns out that *P53* was not an outlier: The burden of mutations in normal skin, in just 74 genes, is about five mutations per megabase. This is only a factor of 10 less than seen in squamous cell carcinoma of the skin and is comparable to the mutation burden seen in cancers of the breast or head and neck. Most mutations in the biopsies have the character-

istic signatures of exposure to ultraviolet light (cytosine-to-thymine changes at adjacent pyrimidine residues), and include nonsignature guanine-to-thymine changes that are caused by many mutagens, including ultraviolet light (7). Copy number changes are also present. To determine whether these mutations were conferring a selective advantage in normal skin, Martincorena *et al.* used the ratio of mutant sequence reads that do or do not alter the amino acid. Six genes—*NOTCH1*, 2, and 3, *P53*, *FAT1*, and *RBM10*—had an excess of nonsilent mutations, indicating that these genes had been selected for and are therefore presumptive drivers. The first five are known drivers for keratinocyte tumors and the last is mutated in other cancers. A seventh gene, *FGFR3*, had recurrent activating mutations. In total, one-quarter of middle-aged skin contains a mutation in one of these drivers.

**“...sun-exposed normal skin is already a polyclonal quilt of driver mutations subjected to selection—a field of preprocancers, as it were—that nevertheless functions as a skin.”**

Whirring in the background of the results is Martincorena *et al.*'s bioinformatics effort to establish gold standard protocols for identifying valid mutations present at low levels. This is critical because a single-copy mutation present in 1% of the cells and sequenced to 500× depth provides on average only 2.5 reads. The supplemental material is an operating manual for identifying rare mutations.

How did our eyelids get so many mutations? In DNA isolated from a homogenized biopsy, a mutant sequencing read could be frequent because its gene has mutation “hotspots” (highly susceptible to mutation) or its cell had a “mutator phenotype” (a mutation that increases the mutation rate at other loci) (8); either route would allow scattered cells to acquire the same mutation

Departments of Therapeutic Radiology and Dermatology, Yale School of Medicine, New Haven, CT, USA.  
E-mail: douglas.brash@yale.edu