

Single-Cell Genomics Dives into the Diversity of Marine Microbes

Researchers use single-cell sequencing on Illumina systems to uncover hidden diversity and biogeographic variability.

Introduction

The single cell is the basis of biology, holding a treasure trove of genetic information within its walls. Yet, the ability to isolate and sequence DNA from a single cell has proven impossible until recently. To overcome this challenge, researchers resorted to performing studies with large cell populations, recognizing that the data obtained doesn't reflect the subtle differences among cells. The field of single-cell genomics finally came into its own less than a decade ago, the beneficiary of new approaches for isolating, extracting, amplifying, and sequencing single-cell DNA. Its growth has paralleled advances in next-generation sequencing (NGS) technologies.

In 2009, the Bigelow Laboratory for Ocean Sciences established the Single Cell Genomics Center (SCGC)¹, the first facility of its kind. "The goal of SCGC is to make single-cell genomics more accessible to the broad scientific community," said Ramunas Stepanauskas, Ph.D., a Senior Research Scientist at Bigelow and SCGC Director. "Single-cell genomics offers vast new opportunities for biology studies ranging from marine microorganisms to cancer and other human diseases."

One of the benefits of performing single-cell genomics in microbiology is that it yields cultivation-independent data, whole-genome data. "Over 99% of the microbes from most environments cannot be cultivated," Dr. Stepanauskas said. "We can't study them using standard microbiology tools, so looking at their genomic blueprints is the most informative way to understand what they are, what capabilities they have, what their roles are in the environment, and what applications we might develop from this enormous biological resource that surrounds us."



Ramunas Stepanauskas, Ph.D. is a Senior Research Scientist at the Bigelow Laboratory for Ocean Sciences and the Director of its Single Cell Genomics Center (SCGC).

In a recent study, Dr. Stepanauskas and his team used large-scale single-cell genomic sequencing to gain insight into bacteria inhabiting the surface layers of the ocean². In the process, they discovered global temperature and latitude patterns of surface ocean bacterioplankton, microbes that have a major impact on the global cycling of carbon, nitrogen, and other elements. They also found much greater genetic variability among uncultured marine microbes than what has been assumed from earlier studies, which relied solely on sequencing the 16S rRNA marker gene.

iCommunity spoke with Dr. Stepanauskas to learn more about single-cell genomics and his marine microorganism research.

Q: What is single-cell genomics and why is it important?

Ramunas Stepanauskas (RS): Cells are as fundamental to biology as atoms are to physics. All life that we know of is composed of cells, while individual components of cells are lifeless on their own. The vast majority of life forms on Earth are microscopic and consist of single cells. Even the human body grows out of a single zygote cell, with all the required biological information encoded in its single genome. For this simple reason, single-cell genomics enables us to address many fundamental questions in biology that we couldn't address until very recently.

For example, researchers studying the human nervous system have identified genomic differences between brain cells that are essential for brain development³. Cancer studies are also starting to take advantage of the single-cell genomics technology⁴. We may find that every human cell is a little different, and those differences might be significant both for the normal function of our body and for the prevention of certain diseases.

Single-cell genomics is particularly important in microbiology, which is my specialty area. It allows us, for the first time, to analyze the metabolic potential and evolutionary histories of the so-called "microbial dark matter" - those 99% of microbial types that constitute the majority of biomass and biodiversity on our planet but have not been amenable to studies using traditional microbiology methods.

Q: Can single-cell genomics identify novel microbes as well?

RS: You can take a drop of seawater or a square centimeter of your skin and you can probably find novel microbes using cultureindependent tools, such as single-cell genomics. However, the identification of novel microbes is not only a technical challenge, it is also a conceptual one. At the moment, there is no commonly accepted, theory-based definition of bacterial or archaea species. We cannot apply the standard species definitions or population genetics and speciation concepts because bacteria and archaea evolve differently from most eukaryotes. They don't undergo meiosis, and their gene exchange is decoupled from reproduction. With single-cell genomics, we're generating data to develop these concepts. Our recent study is an example of those efforts.

Q: Where did you obtain ocean water samples for this study?

RS: We obtained surface ocean samples from the Gulf of Maine, the Mediterranean Sea, and the subtropical gyres of the North Pacific and South Atlantic oceans. The Gulf of Maine represents a temperate, relatively nutrient-rich, coastal environment. The samples from the North Pacific and South Atlantic represent two of the four main subtropical gyres in the world, which are extremely nutrient-poor, sometimes referred to as "marine deserts". The sample from the Mediterranean Sea represents an intermediate environment.

"We may find that every human cell is a little different, and those differences might be significant both for the normal function of our body and for the prevention of certain diseases."

Q: How did you isolate and sequence single-cell genomes from the ocean bacterioplankton in those samples?

Using fluorescence-activated cell sorting (FACS), we deposited individual microbe cells into 384-well microplates. After sorting, these cells were lysed to expose the DNA. We then used multiple displacement amplification to produce single-amplified genomes (SAGs).

We performed PCR on a subsample of that amplified DNA, amplifying and sequencing the ribosomal gene. This gave us a rough taxonomic assignment of the cell. Based on that information, we selected SAGs for genomic sequencing. Of the several thousand SAGs generated, we selected 56 that represented those abundant groups of surface ocean bacteria that had few or no cultured representatives. The Joint Genome Institute performed the sequencing, primarily using the HiSeq[®] 2000 system. On average, 55% of the genome was recovered for each analyzed cell.

Q: How did the SAGs differ from existing bacterioplankton cultures?

RS: We compared our 56 SAGs to the 101 marine bacterioplankton cultures that have been sequenced to date. Our study demonstrated that, with a few exceptions, the existing cultures consisted of copiotrophs, i.e. bacteria that are adapted to nutrient-rich environments. We found that, compared with existing cultures, natural bacterioplankton have smaller genomes (average about 2–3 Mb),

fewer gene duplications, and are depleted in guanine and cytosine, noncoding nucleotides, and genes encoding transcription, signal transduction, and noncytoplasmic proteins.

Q: Does this mean that low GC content is indicative of oligotrophs?

RS: It's still a controversial topic. Oligotrophy seems to play a role in it, but there could be other explanations as well. Guanine and cytosine (GC) contain more nitrogen than adenosine and thymine (AT), so the lower the GC content the lower the nitrogen requirement. It might be that a low GC content is an adaptation for surviving in a nitrogenlimited environment.

There are also other possibilities. GC to AT mutations occur more frequently than AT to GC mutations. If a microbe doesn't have effective repair mechanisms or if the selective pressure is not strong enough, the shift toward low GC will happen spontaneously and might not be due to oligotrophic conditions.

Q: How did you come to the conclusion that GC content rather than protein composition is the primary adaptive trait of bacterial plankton?

RS: We compared the amino acid content of the SAGs and microbial cultures, looking for differences in chemical characteristics of the amino acids. There was no significant difference in pK, nitrogen, or carbon content, or in the energy required to produce these amino acids.

However, we found a higher frequency of amino acids that require less GC in their codons in single-cell genomes than in cultures. That's evidence that the GC content is a primary feature, not the result of selection for a particular amino acid. We also found that the difference in GC content between uncultured and cultured marine microbes was consistent in coding as well as noncoding genome regions. If selective pressure existed for particular amino acids, we wouldn't have seen any difference in GC content in noncoding regions.

Q: Why would a small genome size and fewer gene duplications be an advantage for oligotrophs?

RS: It's just cheaper. Ecology is very much like economics, with "market" rules being the rules of life in the wild. If you can survive and succeed using fewer resources then you will outcompete others.

Small genome size enables the cell to grow and reproduce with fewer nutrients. With a smaller genome, a cell doesn't need as much carbon, nitrogen, and phosphorous to build itself. It can divide sooner, compared to an organism that has a larger genome.

Q: What are the disadvantages of having a small genome size?

RS: This is not entirely clear, but the main disadvantage may be that a cell with a small genome size is adapted to a very specific environmental condition. As these conditions change, the cell may not be competitive anymore. For example, our study indicates that each type of a marine microorganism thrives in a very particular temperature range, with some present only in tropics, others only in temperate regions of the world.

Q: How did SAG data differ from the community small subunit (SSU) rRNA gene sequencing?

RS: Sequencing of SSU rRNA genes has been used as a key tool in cultivation-independent analyses and comparisons of microbial diversity in various environments. In our study we introduced a complementary approach: we used SAGs as references to recruit the existing metagenomic data sets. This provided a significantly improved analytical sensitivity and demonstrated that microorganisms with highly similar SSU rRNA genes have significant genomic and biogeographic variability.

Q: How did the genomic signatures compare?

RS: There are consistent global patterns, with the distribution of ocean microbes based on temperature. We find the same microbes in tropical regions no matter where in the tropical areas we look—Indian, Atlantic, or Pacific oceans. Even more intriguing, the same is true for temperate regions. We find the same microbes off of Maine as we do along the European coast and even Tasmania. That means that the dispersal of marine microbes is not limiting their global distribution.

Q: How did sequencing on the HiSeq system enable this study?

RS: Illumina NGS technology is really the best technology for singlecell genomics right now. The error rate is lower than most other technologies. The paired-end capability provides the capacity to analyze fairly long fragments. We usually need thousand-fold coverage of the genome, and Illumina offers affordable sequencing at this depth. Ultra-deep sequencing is needed because of the uneven amplification of the original, single DNA molecule. Some regions of the genome get over-amplified by a thousand times when compared to other regions. If we want to recover these under-amplified regions, we have to perform very deep sequencing. The combination of paired-end capability, low error rate, and low cost per base makes Illumina sequencing most attractive.

Q: Could you use single-cell genomics to track global warming?

RS: I think it will be an essential tool in many microbial ecology studies, including global climate change. The capacity to analyze the microbial "uncultured majority" will provide the necessary data to study ecosystems that are evolving in response to changes in global temperature. Today, we can start making predictions by looking at the current distribution of marine microbes. We can predict what changes we may expect and those predictions could be tested experimentally and then monitored in the field using single-cell genomics and other research tools.

Q: Have you studied samples from other parts of the ocean?

RS: We are working on a global survey of the genomic diversity of marine microbes, in particular the deep, dark water masses, which constitute 90% of the global ocean and are poorly understood. The microbial composition is very different when you go below the sunlit surface layer, which is typically less than 200 meters deep, while the average depth of the ocean is around 4,000 meters. We are accumulating information about the global differences in the genomic content and metabolic properties of microbes from different water masses, hydrothermal vents, and even deep inside marine sediments and the oceanic crust. MiSeq® and HiSeq® systems are being used to generate data for this project.

Q: What is the next step in your research?

RS: Other than exploring marine microbial diversity, there are many exciting directions that we plan taking. A technical development that is waiting to happen is the ability to perform single-cell whole transcriptome sequencing on environmental microorganisms. It would allow us to study which genes are expressed by various uncultured microorganisms under *in situ* and experimental conditions. Such research would be a next step in understanding the role of environmental microorganisms in the environment and their responses to perturbations, such as the global climate change and ocean acidification.

I'd also like to establish collaborations with industrial partners to take advantage of the vast genomic resources that we are discovering now. The data we are generating could be very useful for bioenergy production or in pharmaceutical discovery.

"The combination of paired-end capability, low error rate, and low cost per base makes Illumina sequencing most attractive."

Q: What types of genomic information could be uncovered from marine microorganisms that could benefit bioenergy research?

RS: One example can be found in a study that we published in 2012, where we employed novel single-cell genomics and single-cell physiology probing techniques to identify and sequence those marine and freshwater microorganisms that degrade laminarin and xylan, two of the most abundant storage and structural polysaccharides in nature⁵. To our great surprise, we found that a few, rare types of Verrucomicrobia make a considerable contribution to polysaccharide degradation in both studied environments. Single-cell genomic sequencing confirmed enrichment of these poorly understood, uncultured microorganisms, in genes encoding a wide spectrum of glycoside hydrolases, sulfatases, peptidases, carbohydrate lyases, and esterases. Remarkably, this enrichment was on average higher than in the sequenced representatives of Bacteroidetes, which are frequently regarded as highly efficient biopolymer degraders.

Degraders of complex organic compounds that are effective under diverse environmental conditions are of great interest to bioenergy, bioremediation, and other industries. Our study sheds light on the ecological roles of uncultured Verrucomicrobia and suggest specific taxa as promising bioprospecting targets. The employed method offers a powerful tool to rapidly identify and recover discrete genomes of active players in macromolecule degradation in complex microbial communities, without the need for cultivation.

Q: What could these studies uncover that could enable pharmaceutical development?

RS: Single-cell genomics uncovers near-complete genomes and intact metabolic pathways from uncultured microorganisms, providing unique access to over 99% of genetic information on our planet. When analyzing single-cell genomes of uncultured bacteria and archaea, we often find pathways of potential interest in novel drug discovery, such as polyketide or antibiotic synthesis. The question is how to bring these discoveries to the point where they're interesting for industry. We need to build bridges between cutting-edge basic research and industry to develop the commercial value of these promising findings.

References

- 1. https://scgc.bigelow.org
- Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, et al. (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A 110: 11463-11468.
- Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, et al. (2010) L1 retrotransposition in neurons is modulated by MeCP2. Nature 468:443-446
- Navin N, Kendall J, Troge J, Andrews P, Rodgers L, et al. (2011) Tumour evolution inferred by single-cell sequencing. Nature 472: 90-94.
- Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PSG, et al. (2012) Capturing single-cell genomes of active polysaccharide degraders: An unexpected contribution of Verrucomicrobia. PLoS ONE 7:e35314.

Illumina, Inc. • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2013 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 1270-2013-008 Current as of 12 November 2013

