

Sequencing a mouthful of microbes

The MiSeq™ System is enabling the Saca la Lengua project research team to sequence thousands of samples to identify the variety of bacteria and fungi that live in the human mouth

Introduction

Most of us were taught not to stick out our tongues in polite company, but this is exactly what biologist Toni Gabaldón, PhD asked thousands of Spanish teenagers to do as part of the Saca la Lengua project.¹ Spanish for “Stick out your tongue,” the goal of Saca la Lengua is to survey the oral microbiome of 15-year-olds from across Spain to learn more about the thousands of different microbes living and thriving in our mouths. Dr. Gabaldón and his colleagues at the Centre for Genomic Regulation (CRG) in Barcelona, are using the MiSeq System to sequence more than 1600 oral samples provided by high school students throughout Spain.

“Saca la Lengua was a good project to showcase the science that we do at CRG,” Dr. Gabaldón said. “Microbiome research is an emerging field that’s expanding quickly. New discoveries are being made every day linking the microbiome and human disease, and how our diet and lifestyles might influence the microbiome. For these discoveries, you need advanced sequencing technology and bioinformatics tools to analyze the data. Thanks to technology developments such as next-generation sequencing (NGS) systems such as the MiSeq and HiSeq™ Systems, we can now assess the diversity of microbes that live in and on our bodies faster and less expensively. That was not possible until a few years ago.”

Saca la Lengua is more than a straight microbial genetics project. It combines citizen science with community outreach to spark the next generation’s interest in science. While data analysis isn’t yet complete, Dr. Gabaldón and his colleagues have already begun making important and surprising discoveries about the oral microbiome.

Studying the fungal microbiome

In the past few years, microbiome research has become a hot topic. However, not all aspects of this new research area are flourishing. Research studies of the bacterial microbiome of the gut and skin have dominated the field. Fungal microbiome studies have lagged, due to the difficulty in culturing fungi and the resulting lack of robust genome databases. They also often comprise a small percentage of the overall microbiome. While their concentrations might be small, the presence of fungi in our oral microbiome can have a profound impact.

“Think of the mouth microbiome as an ecosystem just like that of the African savannah,” Dr. Gabaldón said. “There might be very few lions compared to zebras, but if you remove the lions, you alter the entire ecosystem. Fungi play a similar role. Some fungi

produce antimicrobial compounds; others produce compounds that facilitate bacterial colonization. Fungi consume oxygen, change pH, and otherwise alter the ecosystem in the mouth. When the bacteria *Streptococcus mutans*, one of the species that causes dental caries, grows in concert with the yeast *Candida albicans*, a positive feedback loop is created that causes cavities to grow larger and faster.”

Fungal genomics is more difficult to study than bacterial genomics. To sequence the bacterial genome, scientists have developed universal markers that amplify given regions, with these markers working across many different bacterial species in a similar way. In contrast, it’s hard to break open the thick fungal cell walls to extract and sequence DNA. Existing primers and other tools also tend not to work as well in fungi as they do in bacteria.

As a result, there is almost no information available about the variety of fungi that live in the human mouth. “The few studies that are available are based on small sample studies of people with certain diseases,” Dr. Gabaldón stated. “There are no studies in young populations. Saca la Lengua project provided us with a valuable opportunity to fill this important information gap about the fungal microbiome.”

Saca la Lengua hits the road

Because so little is known about fungi in the oral microbiome, Dr. Gabaldón suggested that CRG Director Luis Serrano, PhD, and the sequencing core at the Centre for Genomic Regulation include these organisms in the Saca la Lengua project. They readily agreed. With funding from La Caixa Bank, Dr. Gabaldón and his colleagues put out a call to teachers in high schools



Dr. Toni Gabaldón is Group Leader, Comparative Genomics and Bioinformatics at the Centre for Genomic Regulation (CRG) in Barcelona, Spain.

throughout Spain. “They focused on 15-year-olds in particular because this is the age when many Spanish teens are deciding whether to focus on science or another subject for their high school degree. Even if participating in the project didn’t encourage some of these teens to pursue science, it would improve their scientific literacy.

Focusing on such a narrow age range had an additional benefit: it made it easier for the researchers to compare differences in the oral microbiome. “These young people aren’t recruited from hospitals, but from high schools,” Dr. Gabaldón said. “Their incidence of diseases such as asthma, reflect what’s found in the normal population. We want to describe the microbiome in this population and how it varies with different factors.”

“The MiSeq System provided us with a winning combination of high-throughput sequencing and long reads.”

“From a research perspective, the Saca la Lengua project has obtained an excellent set of normal samples,” Dr. Gabaldón added. “They’re all from the same regional area in an age-controlled population. If you look at similar studies, you won’t find one with as many samples.”

Besides providing a sample of their mouth microbes, the more than 1600 Spanish teens also answered a series of questions about their diets, hygiene, and lifestyles. The researchers want to compare the effects of these variables on the microbiome, but also the effects of the family’s socioeconomic status, and whether the adolescents live in an urban or rural area.

MiSeq System becomes a sequencing workhorse

The CRG team relied on the MiSeq System to perform the sequencing work. They waited until they had collected all the samples, before performing DNA extractions. They then pooled barcoded samples in the same MiSeq flow cell lane to optimize the number of sequencing runs and the amount of information obtained per sample. Rather than using a whole-genome shotgun approach, the Saca la Lengua team used a targeted approach, amplifying only the region that encodes ribosomal RNA. In bacteria, they looked at 16S rRNA and in fungi, 18S rRNA. These regions provide valuable information about the identity of microbial species. “The MiSeq System provided us with a winning combination of high-throughput sequencing and long reads,” Dr. Gabaldón said. “We performed previous studies with the Roche 454 and the throughput wasn’t even close to what we obtained with the MiSeq System. The reads were also long enough to overlap the DNA region between the two primers flanking the ribosomal RNA sequence, enabling amplification of the entire region of interest.”

The primers used by Dr. Gabaldón’s team to target the ribosomal RNA gene also include Illumina functional adapters, providing a seamless transition between library preparation and sequencing. The completed libraries were then sent directly to the CRG core lab for sequencing on the MiSeq System. “We’re interested in sharing what we’ve learned,” Dr. Gabaldón added. “The protocol we used is available to scientists who are interested in performing this type of study.”

In addition to the MiSeq System, the team uses the HiSeq System for resequencing fungal genomes and for *de novo* sequencing. “From the genomes, we prepare paired-end and mate-pair libraries and also perform RNA-Seq,” Dr. Gabaldón said.

Where are the fungi?

Shortly after they began sequencing, Dr. Gabaldón and his team realized that many of their samples didn’t show any signs of fungi. They cultured some of the samples and double-checked the sequencing results against samples from CRG staff who were over the age of 30. It verified that low concentrations, or the complete absence of fungi, was characteristic of the oral microbiomes of this teenage population. “We were not expecting this result and it might have some implications,” said Dr. Gabaldón. “It might be that fungi colonize our oral microbiomes later in life. We’re not sure if there are correlations with dietary habits. It’s too early to say.”

“With Illumina sequencing systems, I can perform whole genome sequencing, or by altering a few conditions, perform RNA-Seq and obtain a large amount of data to answer a genomics question.”

Aside from the surprising absence of fungi, sequencing of the samples has been easy and straightforward. The data analysis portion of the project, has been more complicated. The Saca la Lengua team obtained millions of reads per sample, representing the rRNA gene region in different bacterial species. That amount of data presents a challenge in determining which bacterial or fungal species or genera are present in each sample. Dr. Gabaldón used an analysis pipeline that mapped the MiSeq reads to a database and used a scoring system that measured the amount of alignment and the similarity of sequences. With these scores, the researchers could say whether they had *Streptococcus mutans* or another *Streptococcus* species.

A better understanding of the oral microbiome

“We are very happy with the quality of the MiSeq data,” Dr. Gabaldón said. “Any problems that we’ve had have been generally due to the quality of the DNA.” The MiSeq System has also opened a door to understanding the oral microbiome. “With

Illumina sequencing systems, I can perform whole-genome sequencing, or by altering a few conditions, perform RNA-Seq and obtain a large amount of data to answer a genomics question,” Dr. Gabaldón added. “A few years ago, that was impossible. I had to rely on publicly available sequencing data from large consortia for certain genomes. With the Illumina systems, I can be proactive and if I need a certain genome, I can sequence it myself.”

Data analysis has just begun and Dr. Gabaldón doesn’t yet have information on the bacterial and fungal species they found and what amounts, or how the oral microbiome might be related to lifestyle and health variables. That hasn’t stopped other researchers from taking an interest in the project. “I’ve received interest from researchers to expand the project and initiate studies in their countries,” Dr. Gabaldón said.

Connecting with the citizenry

An unexpected benefit of the Saca la Lengua project has been the ability to connect with teachers and students. “This is the first time I have been involved in a project that has brought me so close to the citizens of Spain,” Dr. Gabaldón said. “I’ve met with teachers and students participating in the project, discussing it with them and answering their questions. Giving a talk in front of an audience of 15-year-old students is harder than talking about your research with a group of scientists at a meeting. You need to make a significant effort to connect with them.”

Moving forward

Dr. Gabaldón hopes that the sequencing data and metagenomics studies of these specimens will enable him and other researchers to learn more about the normal, healthy oral microbiome, and how assorted variables can alter the make-up of mouth microbes. These insights, in turn, might provide more information about various chronic diseases that affect teens and adults. “As we learn more, we are realizing the microbiome is connected with diseases such as Crohn’s disease, and some diseases that you might not even think of, such as arteriosclerosis,” Dr. Gabaldón said. “In principle, it seems odd that there might be a connection. Yet, every time you brush your teeth there are some cells that enter the blood stream. Some of them are able to trigger a significant immune reaction that might create granules in the blood that could cause veins to collapse, for example. It all depends on which bacterial species you have in the mouth. We are learning that the microbiome is a factor that influences our health status and might be a factor in some important diseases.”

He also points out that the oral microbiome has advantages with sample collection, because a saliva sample is far easier to obtain than a biopsy. “The more scientists understand the oral microbiome, the more research opportunities open up to improve human health,” Dr. Gabaldón added.

Learn more about the Illumina system mentioned in this article:

MiSeq System, www.illumina.com/systems/miseq.html

References

1. Saca la Lengua. www.sacalalengua.org. Accessed March 31, 2016.

