Personal Science Guide to the Microbiome

2018-06-10

Contents

Intro

Science is about curiosity, a passion for truth, and a never-ending process of systematic experimentation in pursuit of an understanding of the world around us. It's open to everyone and can be applied to any situation, including situations that affect you right now. It's *Personal Science* when you use the scientific method to discover important insights about the wellness and performance of yourself and those around you.

This book will introduce you to one exciting area of personal science: understanding your own microbiome. Our companion web site lets you upload your own data and compare it to others to make your own discoveries.

Most of the examples are based on over 600 near-daily samples I took of my own microbiome over a three year period. Inspired by an experiment conducted at MIT[1], during most of that time I also carefully tracked the food I ate, my sleep, and other variables like activity or location. Most of my near-daily samples were of my gut, but I also regularly tested my skin, nose, and mouth. Since I'm generally healthy, I didn't have a specific goal in mind other than to try to understand better what these microbes are doing, so many of my tests were taken while undergoing simple experiments, like eating a specific type of food or visiting a new location. While not necessarily up to the rigorous standards of a formal scientific trial, these "n of 1" studies on myself helped me discover several new interesting facts about my own microbiome, many of which appear to contradict other published studies. In addition, hundreds of people sent me their own test results, letting me compare many different microbiomes. And of course, I also followed the latest developments in scientific publications and the general press as I eagerly tried to learn more.

This book tells you what I learned.

[1] David et al. (2014)

Chapter 1

Getting Started

So you tested your microbiome and now you want to know what it all means. Later we'll go into much more detail about the sampling process itself: differences between labs, how to get the best sample, how long it will take, etc. but let's start by assuming you've gone through all that. What will you learn?

1.1 Using the uBiome web interface

You just received an email that uBiome has finished processing your sample. Now what do you do?

The first stop is the uBiome Explorer web dashboard, where you'll see an overview of your results. The dashboard is intended to be self-explanatory, so you should feel free to click around to try the various options. The beginning page will look something like this:

If you only have one sample, your page will look something like Figure 1.1.

The colorful pie chart there is your first high-level glimpse of the bacteria found in your sample. Hover over it with your mouse (or tap with your finger if you're on a phone or iPad), to see a pop-up view of the approximate percentages.

You're seeing a *phyla*-based view of your microbiome. As we'll learn in more detail later, this is a very high level summary, a bit like how a Martian might see life on earth from far in outer space. Keep in mind that by "very high level", we mean *very* high, the equivalent of seeing life only at the level of "vertebrate" and "invertebrate". The incomprehensible diversity of bacterial life means that at this level, you don't even bother separating fish from one another or for that matter, from land animals: just lump all of them together. At the phyla level, mammals, reptiles, fish – they're all the same. Similarly, when you see the phylum *Firmicutes*, for example, think of it as having as encompassing as much variation as all the vertebrates, from those on land, in the ocean, big, small, harmless, dangerous and everything in between. In other words, this first approximation is pretty crude.

Note that the various phyla are represented as percentages of the whole, so some of them are relegated to tiny slices of the pie. That doesn't necessarily mean they're unimportant, though, because in the world of life, even some tiny populations can exert an outsized influence. A wolf is just one of a gazillion different creatures in Yellowstone National Park, but that is little comfort to a rabbit.

Remember too that so far we're treating each individual organism as if it has the same effect as the others, another oversimplification, just like it would be a mistake to assume that elephants matter more because they're bigger than humans.

All we know at the phylum level is that, proportionately, certain large classes of bacteria are more well-represented than others. That may not sound like much to go on, but as we'll see, scientists have been able to learn quite a bit from tiny clues like this.

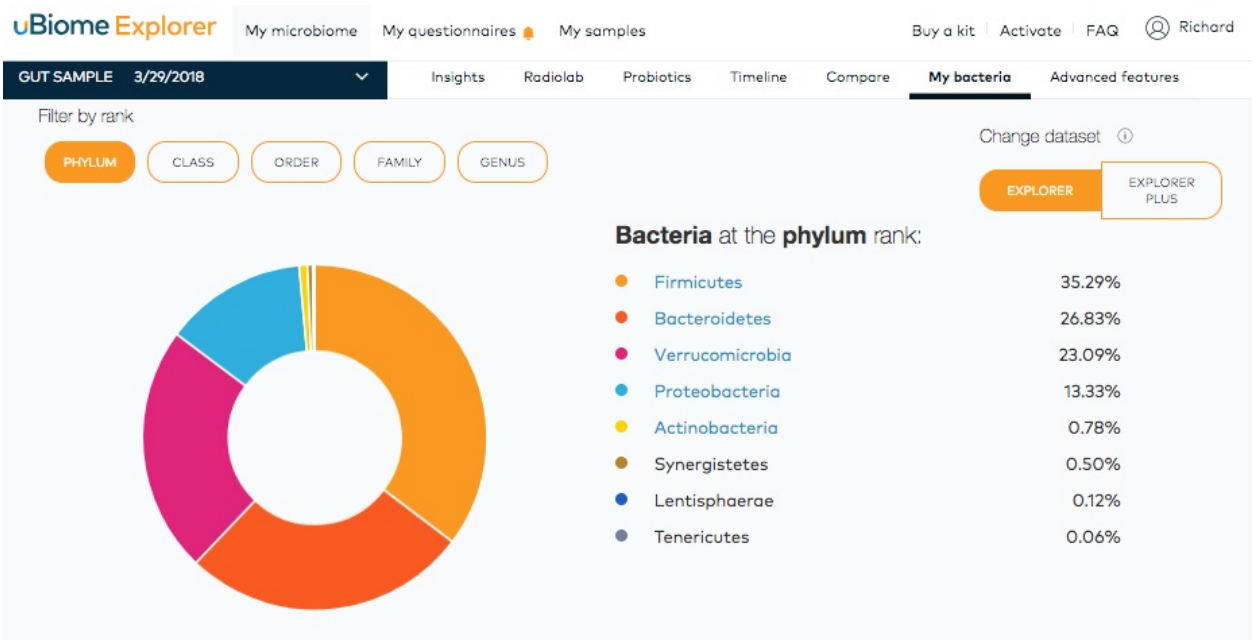


Figure 1.1: uBiome Explorer Insights Overview

Click next on the "Insights" tab (Figure 1.2) which may be more informative for beginners.If this is your first time here, go ahead and breeze through this page, but just ignore most of the conclusions. As we'll learn, scientists have barely scratched the surface of what we will eventually understand about the microbiome, so think of this page as a teaser – a hint of the potential. If your results show something you didn't expect, say on diversity or body weight, don't worry: it almost certainly just means that the Insights pane is wrong about you. Later we'll see how to pull out far more interesting and useful results.

The one part of the Insights page that you *should* take seriously is the section called "most uncommon bacteria". Let's jump to that section now.

1.2 How are you different?

Since so little about the microbiome is well-understood, most of the time it will be hard to interpret your results by themselves. This is where the power of uBiome's citizen science can help, because the web page lets you easily compare your results with other people.

In this case, I can see that my sample has about 64% of the abundance of *Firmicutes* as other people in the uBiome database. You can sort this field by abundance (ratio) and click on the various taxonomic categories to dig deeper into your microbiome and find microbes that are particularly rare in your sample. (Figure 1.4)

For example, I found that I have much less *Prevotella* and *Butyvibrio*, two common microbes than the vast majority of uBiome customers. I also found much higher-than-normal levels of *Deinococcus-Thermus*, an extremophile usually only found in harsh environments, like the impossibly acidic human stomach. Why would I have so many more than other people like me?

One clue is the similarly-high levels of Proteobacteria, which as we just noted is often high in people with gastro-intestinal issues. Maybe I was exposed to something that released a bit of stomach acid, and its extremophile *Deinococcus* into my colon.

Verrucomicrobia is elevated too, though not by much. This is a Phylum that is generally associated with soil microbes, though it's not unusual to find it in the gut as well.

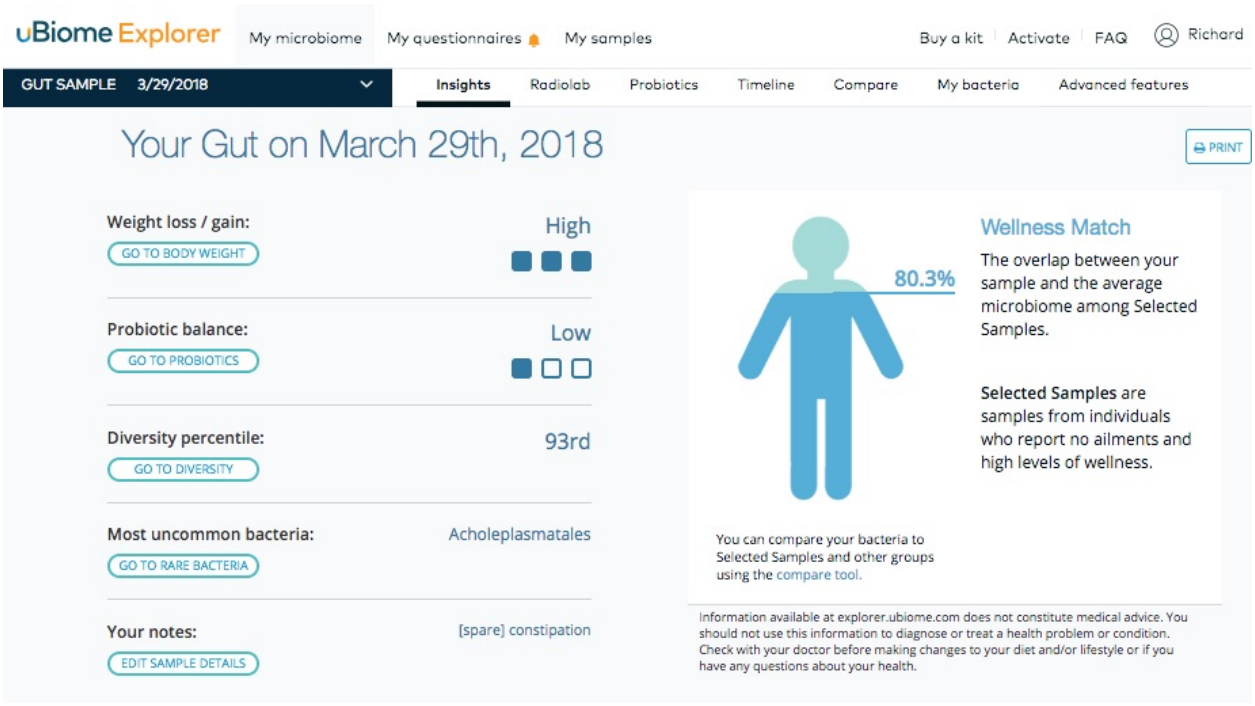


Figure 1.2: The Insights tab provides a high-level view of what uBiome thinks is most actionable about your microbiome.

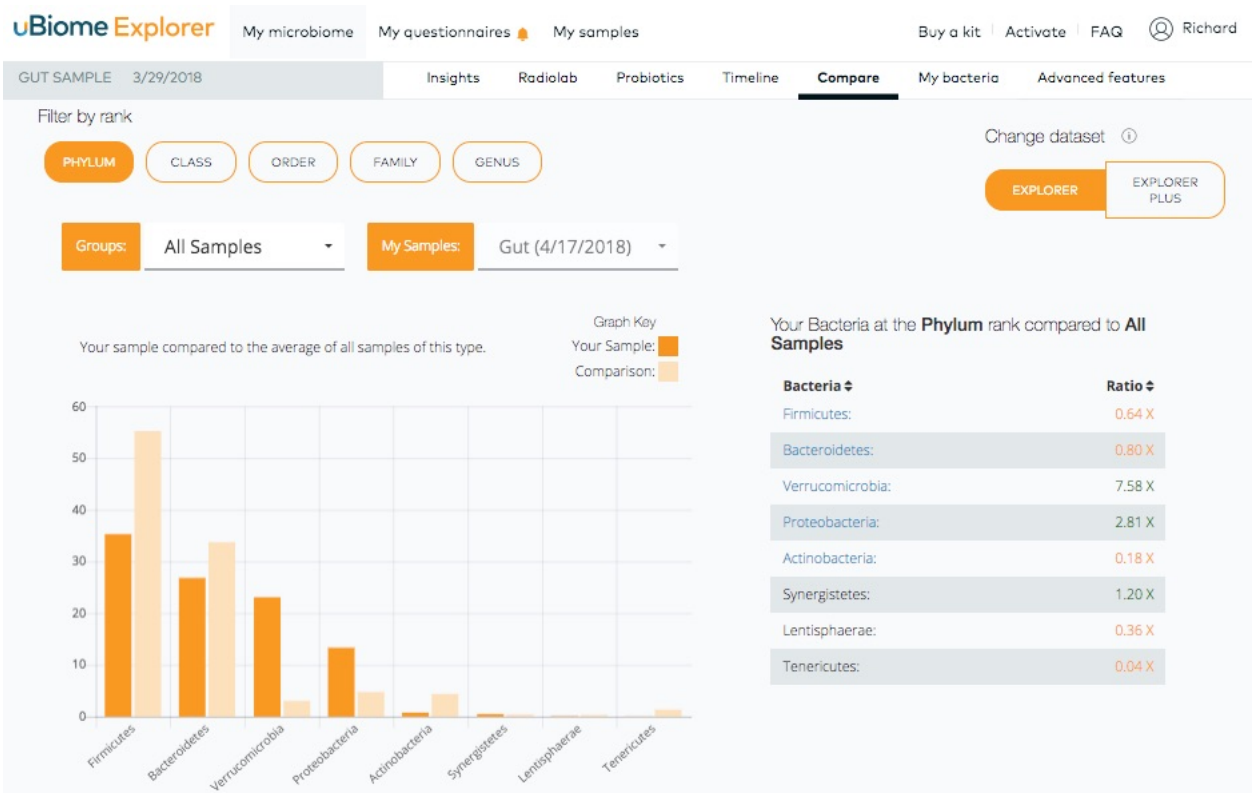


Figure 1.3: The Compare view shows the differences between my sample and others in the uBiome database.

Figure 1.4: My sample shows very low levels of these microbes, compared to the hundreds of thousands of others in the uBiome database.

I looked back at my notes during the period of a few days before this sample, and interestingly I had been doing work in my garden at the time. We have deer and other animals that come near our house...is it possible that I was exposed to some animal excrement, maybe some of which somehow made it into my gut?

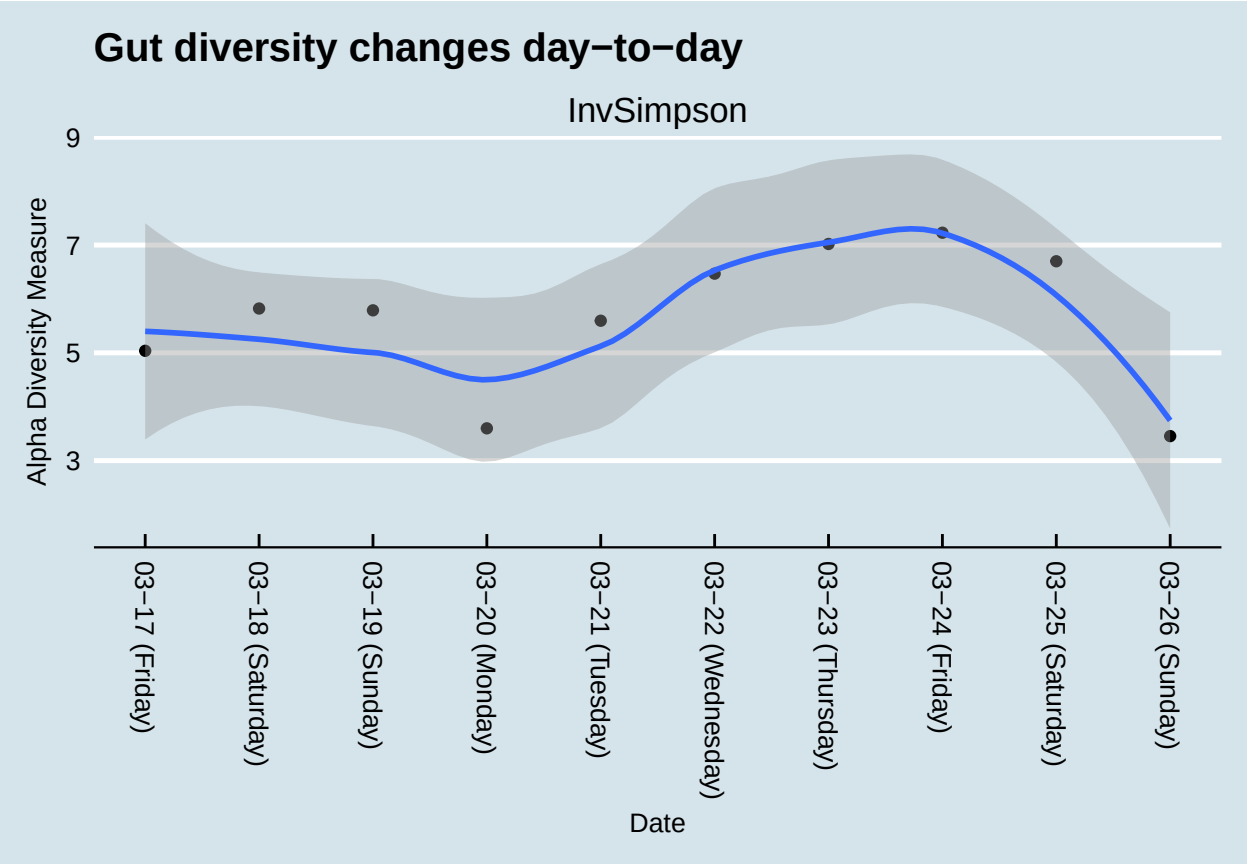
It's just a theory, and perhaps an unlikely one, but I wonder if my body was fighting some type of pathogen as a result of my outdoor activities. Now before we get too excited, keep in mind that I never felt the slightest bit ill during this period. Can I thank my microbiome? Perhaps I was exposed to something nasty, the kind of bug that under other circumstances might have wreaked havoc and left me terribly sick for days. Now I have no way of knowing for sure, but this tidbit makes me wonder if maybe my microbiome acted as a guardian angel, protecting me from something much worse.

Similarly, your own results comparing to others may lead you to clues about aspects of your environment you didn't otherwise suspect.

Incidentally, since this test I've sent dozens of additional samples to uBiome and never saw that *Deinococcus* thing again. Good riddance? or maybe I should be thankful for the protection, and hope that little bits of it may still be present, undetectably in my stomach, waiting to fight another external enemy the next time I come in contact with something in the soil.

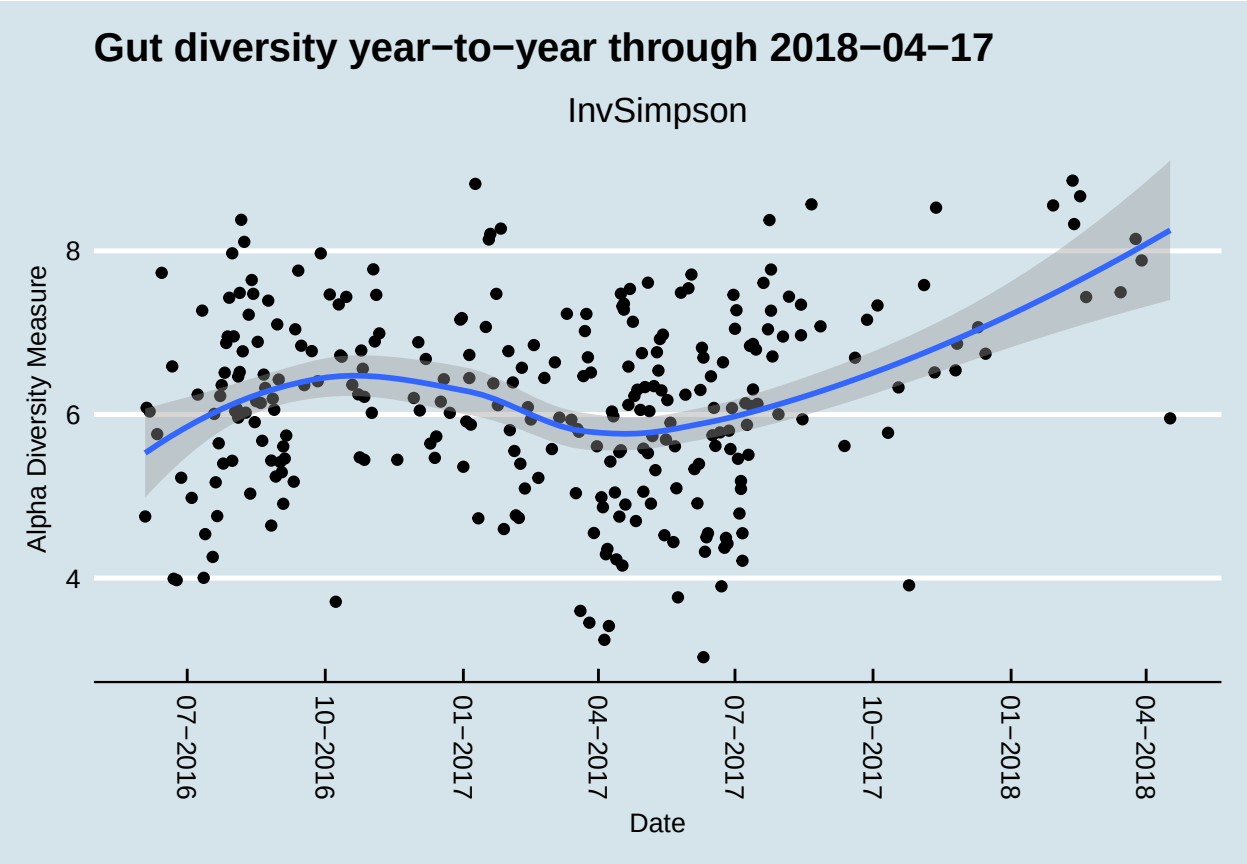
1.3 Diversity

The general consensus is that diversity is good: a greater variety of microbes ensures more resillience against the daily threat of invaders. Many people, after taking just one test, often feel either reassured that their diversity is "good" or disappointed that it's "bad". But I find that day-to-day variability is great enough that it's almost never useful to use a single result. For example, here's my diversity during a typical week:



If I had only tested on Monday, I may have been disappointed with my 0.38 score. Wait another day or two and, with no significant changes in diet, I was up to 0.72 – before plunging to 0.37 by the weekend. Moral: don't take a single result too seriously.

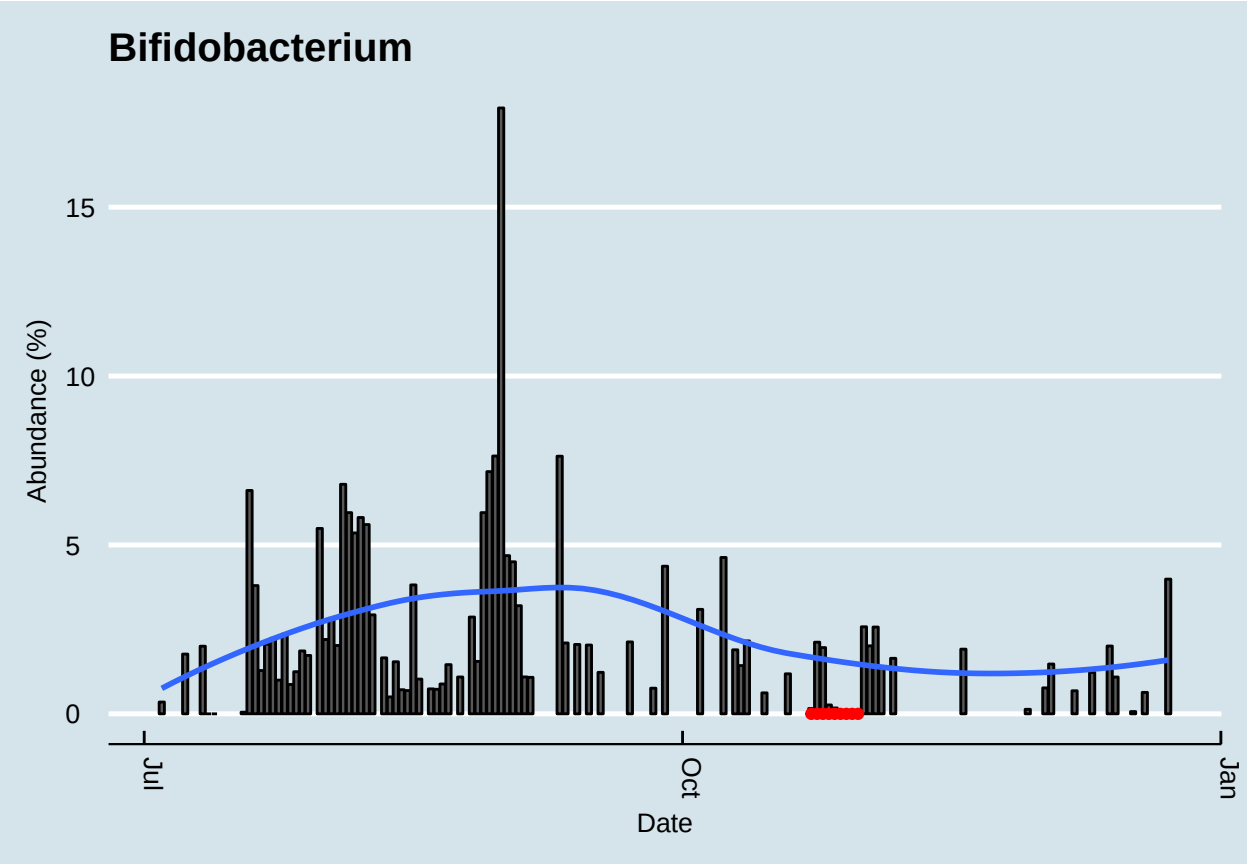
To get a sense of how much diversity can vary over a year:



Although the blue moving-average line shows apparent stability, there are many days that are far above and below the average. Yes, at various times during this period I was eating different types of food, often in a deliberate attempt to influence my microbiome, but believe me: that is not the reason for the wild changes up and down. I also studied the mathematics behind how diversity is measured, hoping to find something more "accurate", but ultimately I concluded that, like many attempts to summarize the microbiome in a single number, the whole concept of diversity is a mirage. Everything depends.

Bifidobacterium is a key component of virtually all popular probiotic supplements, partly because it is so easy to manufacture, but also due to its proven association with sleep and other aspects of health. A six month picture of my levels shows some dramatic ups and downs[1].

[1]Important note: in all of these charts, I don't distinguish between a day when the tested abundance was zero, and a day when I have no reliable test results. Besides the daily variability in the microbiome, the test results themselves often come with wild differences in quality. This is nothing to complain about: the ability to see these microbes in the first place is a true miracle of technology, and even the most careful lab can't get around the myriad of technical issues that come with such a complex process.



Incidentally, the red dots indicate days when I was taking a powerful probiotic supplement that contained *Bifidobacterium*. And that big spike in September? That was during a trip to New Orleans, when I ate a lot of red beans and rice. At least for me, food seems to work better than taking supplements.

Chapter 2

Microbes to Watch

2.1 Phylum

Debunking the Firmicutes/Bacteroides ratio

2.2 Genus

Bifidobacterium and sleep

2.3 Species

Chapter 3

Methods

This chapter will go into more detail about methods, building on the Explore Your Microbiome chapter to show more precisely how I measured myself and how I used the tools needed to build this book.

3.1 The technology for studying microbes

People have been farming the microbes in fermented foods for thousands of years, so when in Pasteur times, scientists first began to cultivate them for experiments, the most obvious way was through the process known as "culturing". Take a sample containing some microbes of interest, and leave them sit in a hospitable environment long enough for them to reproduce in enough quantity to be useful. That's still a common way to study microbes, and that couple-of-day incubation period is one reason you don't get your lab tests back for a few days.

Culturing also has several serious limitations. It only works if the microbes are still living, which rules out many important situations. Many microbes don't culture well or at all outside their native habitat.

Anaerobes are organisms that can't survive in the presence of oxygen, not a problem deep inside the airless gut, but it won't work in a normal lab. While you can take some precautions to preserve the original environment as much as possible – you can set the organisms in a specially-sealed oxygen-free container –the cost and expense rises quickly.

Even if, somehow, you were able to overcome all the other challenges, many (perhaps most) microbes don't grow well unless they are in close proximity to other specific species. *Methanobrevibacter smithii*, for example, which plays a critical role in the efficient digestion of complex sugars, removes hydrogen from its environment, providing a habitat for organisms that don't like hydrogen, like *Firmicutes* and *Bacterodetes.* Plus, it converts all that excess hydrogen to methane, which in turn is needed by yet other organisms. Culturing any of those microbes on their own would be difficult, if not impossible.

But the techniques for uncovering which organisms are where and what they are doing was revolutionized in the first decade of the 2000s by those new-fangled gene sequencers that were so usefully applied to human genes.

3.1.1 The 16S rRNA Gene

Despite the plunging costs of DNA sequencing, the trillions of microbes in your gut still present a formidable chellenge if you intend to sequence them all. Even the humble *E. coli* contains nearly five million DNA letters. There is some commonality between related organisms — humans and chimpanzees, for example, share upwards of 90% of their DNA — but in general it's hard to use the DNA strand itself to measure the

relatedness between two organisms. Understanding the reason for this may help you understand why there is a clever shortcut.

You might think you can measure the relatedness of two organisms by looking at all the DNA in each one and computing the percentage that each shares in common. This would work, but sequencing all those billions of DNA bases takes a lot of time and money, and it would be impractical in a case like the microbiome where you may need to do this for millions of individual organisms.

A service like 23andme is able to cheaply compare individuals of the same species (i.e. Humans) because the generic human genome is already well-mapped and we know that of the 3 billion base pairs, only about 3 million (the SNPs or single-nucleotide polymorphisms) are different between individuals. When you give your spit sample to 23andme, they give you back a subset of your SNPs, only those that have been studied enough to be interesting. SNPs are easy and cheap to find using a "gene-chip", a special semiconductor-like device that can quickly look at 1 million or more pre-determined spots on your DNA. But this is only possible because the map itself already exists, thanks to multi-year effort of the Human Genome Project that finished in the early 2000s. There are no comprehensive gene chips (yet) for bacteria, and certainly not for all the millions of species in nature. And even if there were such chips, bacteria are notorious at adapting and changing to their surrounding environment, exchanging genes with one another, that it just wouldn't be practical to identify enough constant genes to make it worthwhile.

Fortunately, to get an overall picture of the types of microbes in your body, we don't have to sequence every piece of DNA. For our purposes, we just want to know *which* organisms are there, and in what abundance. The precise bits of DNA are important only if they let us know the names of the microbes, and for this we don't need to bother sequencing everything. In fact, most bacterial species differ enough from each other that we need only a few bits of DNA from each in order to tell them apart.

We know that all bacteria are distantly related to one another, and that closely-related species will have more DNA in common with each other. But some of parts of DNA are so important that they stay virtually identical even across entire families of organisms. Remember that DNA describes absolutely *everything* about the organism, including the workings of very low-level cell process. Not just the size or shape, but much more fundamental: how a cell divides, for example, or even how to use the oxygen a cell needs for survival.

Among the most fundamental of all processes is what happens in every cell's ribosome, a special molecule that is core to how a cell converts DNA into proteins. Because all cells create proteins, they also always contain a ribosome and, importantly, they contain the *instructions for how to create a ribosome* in the form of the ribosomal gene. Each cell's DNA includes a gene that precisely encodes every protein, in the exact order that makes up the ribosomal structure. A special enzyme, called DNA polymerase, manufactures new bits of RNA on the fly as it hits portions of the DNA. These bits of RNA, called messenger RNA or mRNA, eventually make their way to ribosomes, which are floating throughout the cell. Upon hitting the ribosome, mRNA is converted into the proteins that make all life possible. If it happens that the mRNA hits upon a segment of DNA that encodes a gene for a ribosome, guess what new molecule is manufactured? A new ribosome!

This ribosomal gene is such a fundamental part of every living organism that very little about the ribosome changes, even after hundreds of millions of years of evolution. Humans and corn plants actually share quite a bit of the ribosome; both are prokaryotes, for one thing, so many of our cellular processes work the same. But bacteria go back even further than humans and corn plants, enough so that the differences aren't so subtle anymore. In fact, the differences are big enough that, with clever selection of the part of the genome to sequence, you can tell the difference between two bacteria in a few hours for a fraction of the cost of running through all the DNA you might find in a microbiome.

The gene that encodes ribosomal RNA (written rRNA) for bacteria consists of about 1500 base pairs total, a tiny fraction of the entire genome, and although it is mostly identical across all bacteria, there are *some* differences, all of which are contained in nine "hypervariable" regions containing even fewer base pairs. These regions, named V1 through V9, are surrounded by strings of base pairs that are constant throughout all bacteria, and can be quickly discovered and amplified by the right DNA primers. The fourth one of these regions, V4, contains only 250 base pairs, and is quickly and easily sequenced on commercially-available

sequencing machines.

When you submit your sample to a lab, the bacterial cells must first be cut into pieces ("lysed", to use the technical term). Sometimes the first part of this process happens at collection time, when you swab a tiny bit of your sample into a vial and stir. The vial contains tiny "beads" that smash into the cell walls as you stir, breaking them apart to spill their contents in an ugly liquid "goo".

The lab is interested only in the DNA inside that goo, so they start by dropping in some carefully-constructed "primers". These are bits of known, synthetically-made DNA that are designed to bind just to the parts of the cell DNA that make ribosomes. In particular, these primers will only find bits of DNA that make the specific, V4 subregion of the ribosome. Primers naturally bind and then break open the DNA at precise locations, cutting out all the segments that match.

Throw this goo into a centrifuge spinning at a carefully controlled, very high speed, and different parts of the goo will rise to different levels, reflecting their molecular weights. One specific part, corresponding to the section programmed to make ribosomes, will rise to a centrifuge level referred to as "16S". Precisely skimming the goo at that spot will give the technician a collection of DNA from just one part of the ribosome of bacteria. The rest of the DNA, millions of letters (base pairs) per bacteria, will not be sequenced and is simply discarded. That's the shortcut. Instead of sequencing millions of base pairs, we need sequence only hundreds.

Once you have a bunch of that 16S ribosomal gene, you know that you are looking exclusively at non-human bacterial and archaeal DNA. It's a tiny subset of all the genetic information in the microbiome, but combined with one more shortcut, it gives a surprisingly accurate look at the overall composition of a sample.

The remaining shortcut is possible thanks to years of research of sequencing the genes in bacteria. Scientists in labs around the world have been faithfully digging up samples of bacteria, and performing whole-gene sequencing on what they find. Although 250 base pairs may seem like a tiny number to differentiate among all possible bacteria on earth, for gut microbiome purpoes we need concern ourselves only with those that are known to inhabit humans. The Human Microbiome Project already identified most of these bacteria – and their 16S gene identifiers – so armed with that as a reference database[1], it is generally possible to unmask a specific microbe with just a sliver of DNA.

It's this two-step combination, 16S "skimming" and a database lookup, that makes it cost effective to study the millions of organisms in your microbiome. You don't have to do a complete gene sequence on every single bacterium; just trust that the tiny subset of DNA in the 16S region is enough to uniquely match something already in the bacterial database.

The alternative – sequence everything in the sample – provides much more accuracy of course, but the 16S approach comes surprisingly close. Careful studies that compare with the "sequence everything" (aka metagenomic) approach show that 16S is still surprisingly close – at least 80% and often much more of the entire microbiome can be categorized accurately, even at the species level.

[1] A very popular one is Greengenes: http://greengenes.secondgenome.com/downloads. Learn about all the big ones here: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3501-4#Fig3

Bibliography

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., Erdman, S. E., and Alm, E. J. (2014). Host lifestyle affects human microbiota on daily timescales. *Genome Biology*, 15(7):R89.