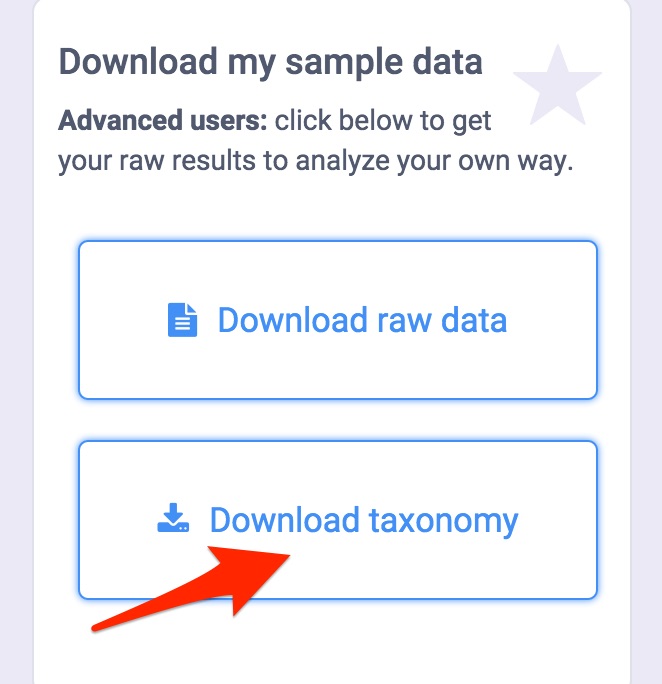
How to analyze your raw uBiome samples

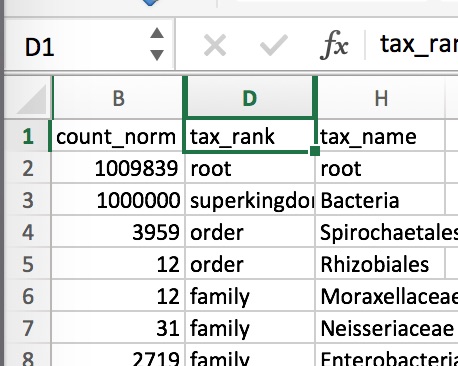
If you only submit one or two samples, the standard uBiome web page (at <http://app.ubiome.com)> offers plenty of information about your microbiome. You can look at the percentage breakdown of different bacteria, compare them with other users or to yourself over time, and dig deeper with descriptions of the most common organisms and what they do. But if you really want to understand your microbiome, uBiome offers much more: full access to all the raw data, literally millions of snippets of genetic information ready to analyze.

I recently wrote a detailed description for the July 2015 issue of O’Reilly’s Biocoder magazine (available as a free download here: <http://www.oreilly.com/biocoder/> ) and I encourage you to read the whole thing for more details, but here’s a short summary. Three steps to get more from your data:

# STEP 1

First, click the “Download taxonomy” button on [the web page for your sample](https://app.ubiome.com/). 

Although it will look like gobbledygook, you can turn this into an Excel spreadsheet easily enough. Select the info on the page and copy/paste it into a site that will convert it automatically into a CSV file of Comma-Separated Values. (I use <http://konklone.io/json/> or <http://www.convertcsv.com/json-to-csv.htm)>. Open the CSV file from Excel and ignore everything but these three columns: tax\_rank, tax\_name and count\_norm.



Now it's a simple matter of running some standard Excel filtering operations on the data. Filter tax\_rank by “phylum”, "species" or whatever other taxonomic rank you care about and then sort the count\_norm field from largest to smallest. The count\_norm numbers correspond to parts per million; divide by 10,000 to convert to percentages of that sample.

By the way, a big bonus awaits you in the taxonomy file that you can’t get from the standard uBiome web page: species information. [Most scientists trust](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2045242/) the 16S rRNA technology down to the genus level, but there is more uncertainty at the species level, so uBiome doesn’t publish it to the web page. Drag it into Excel, though, and you can make up your own mind about whether you trust the species info or not. (And as always, keep in mind that you should never treat uBiome results as medical information; if you're sick, see a doctor.)

# STEP 2

Head to the new [uBiome open source microbiome-tools GitHub page](http://www.ubiomeblog.com/invitation-to-join-the-ubiome-github-repository/) and [download ubiome.py](https://github.com/ubiome-opensource/microbiome-tools/blob/master/ubiome.py). If you have the Python language on your computer you can run this file without installing anything extra. (All Macs come with it built-in; Windows users [download it for free](http://python.org)).

Let’s say your spouse has the sample in a file called Wife1.JSON and yours is in Husband1.JSON. On a Mac, open Terminal and run the following command:

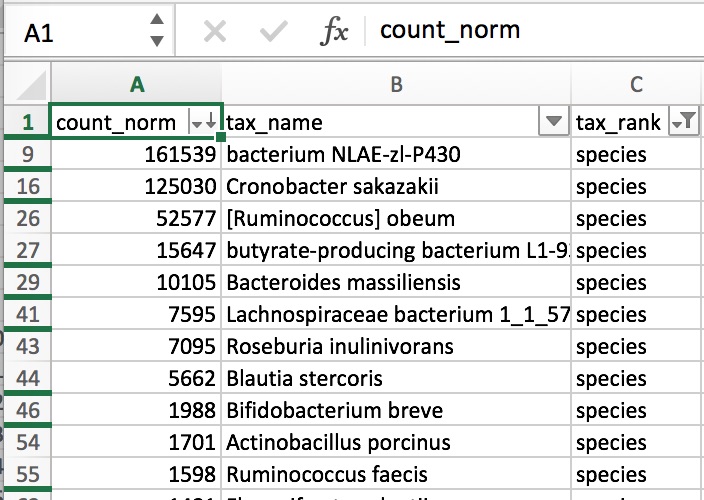
> python ubiome.py –u Husband1.JSON Wife1.JSON \> HusbandUnique.CSV

The new file, HusbandUnique.CSV contains just those organisms that are unique to the Husband1 sample, i.e. are found in the husband’s microbiome and not the wife’s.

Similarly, the following command will give you a file that contains the relative differences between every organism in Husband1 and Wife1:

> python ubiome.py –c Husband1.JSON Wife1.JSON \> HusbandWifeCompare.CSV

Read HusbandWifeCompare.CSV into Excel, sort and filter it, and you'll see something like this:

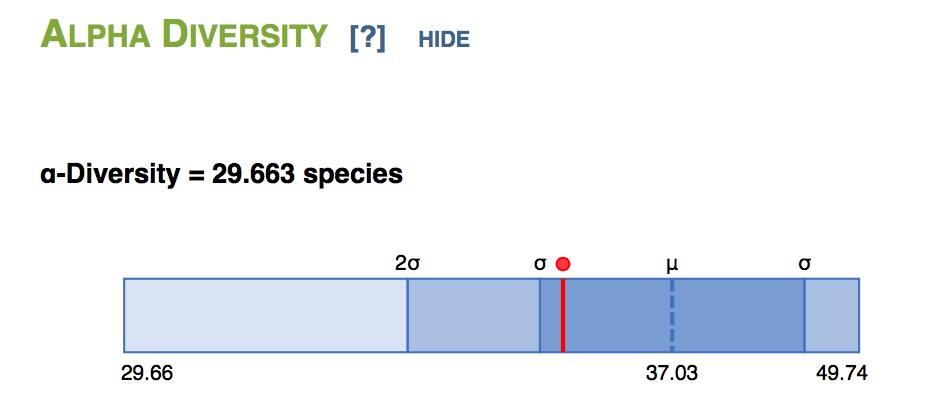


Positive numbers indicate more in the Wife1 sample; negative numbers are more in Husband1. Armed with this information, you can try to understand the reasons behind the differences. Follow [the uBiome Blog](http://www.ubiomeblog.com/category/learning/) for [some](http://www.ubiomeblog.com/one-week-change-in-my-microbiome/) [examples](http://www.ubiomeblog.com/my-microbiome-in-the-jungle/) of how I've done this.

# STEP 3

Finally, if you’re really into serious number crunching, uBiome gives you the raw output from a state-of-the-art Illumina NextSeq 500 in the form of FASTQ files. If you know what that is, you probably already know how to read them, but if not please look at the BioCoder article for an introduction. With a little work, the FASTQ files will let you see precisely which genes were detected in your sample. Since so much of the microbiome is still unexplored, you may find pieces that are missing from the regular uBiome output, so this is your chance to go straight to the underlying genetic information for more.

For example, I was able to compute the following measure of diversity from my most recent sample. It’s a measure [I’ll track](http://blog.richardsprague.com/2015/05/my-gut-diversity-through-time.html) for all of my samples:



# Going Even Further

I’ve barely scratched the surface of what’s possible when you use your raw uBiome data. Please look at the [BioCoder article](http://www.oreilly.com/biocoder) for more step-by-step instructions, and [contact me](http://twitter.com/sprague) if you have other questions. Meanwhile, please upload your own samples to the [uBiome Open Source GitHub](https://github.com/ubiome-opensource) and feel free to compare yours to the ones there, so we can all learn more together.