

**Fwd: FW: How can we download a larger metadata file from the genometrakr website?**

Doster, Enrique &lt;Enrique.Doster@colostate.edu&gt;

Wed 5/29/2019 9:20 AM

**To:** Meera Al Shamisi <alsh0061@umn.edu>; Noelle Noyes <nnoyes@umn.edu> 1 attachments (9 MB)

PDG000000002.1379.metadata.tsv.zip;

Hi Meera and Noelle,

As requested, here is the set of emails I exchanged with Ruth from genometrakr. Let me know if you have any questions, I've been playing around with downloading all the metadata files, but keep running into issues with sftp kicking me out after just a few files. My next plan of action is writing a python scripts that downloads those metadata files in batches. Then we can parse the files to only select the genometrakr isolates we are interested in.

I haven't been able to be productive during my trip but I am confident we'll be able to download the metadata files, even if it's one at a time for each of the isolate groups since there aren't that many for each species.

Thanks!

Enrique

----- Forwarded message -----

From: "Timme, Ruth" &lt;Ruth.Timme@fda.hhs.gov&gt;

Date: Apr 27, 2019 3:55 AM

Subject: FW: How can we download a larger metadata file from the ?

To: "Doster, Enrique" &lt;Enrique.Doster@colostate.edu&gt;

Cc:

Hi Enrique, can you explain a little more about the filtering you need? The metadata file on the FTP site should contain all the salmonella isolates in the Pathogen Detection database. This includes all the GenomeTrakr contributions as well as PulseNet, Public Health England, and other submitters. If you want to filter for GT-only isolates I can help you identify some flags, like the sra\_center or bioproject\_center that you can filter on for isolates submitted by GenomeTrakr. We don't draw a clear distinction in the database for the submitting networks since the data are all collected as part of the same surveillance effort, from the same SOP, QC thresholds, etc.

I'm attaching the sheet I just downloaded so you can see the headings. I did a quick filter for sra\_center = FDA, CFSAN, FDA, CFSAN or FDA/CFSAN and the numbers look about right for what we've submitted for the network (almost 40K isolates).

Best regards,

-Ruth

---

**From:** "Doster, Enrique" <Enrique.Doster@colostate.edu>

**Date:** Friday, April 26, 2019 at 2:52 PM

**To:** "Timme, Ruth" <Ruth.Timme@fda.hhs.gov>, GenomeTrakr <GenomeTrakr@fda.hhs.gov>

**Subject:** Re: How can we download a larger metadata file from the ?

Hi Ruth,

Thank you so much for the quick response.

Your answer was close to what I needed. I was hoping there was a way to download a metadata file for example, that included all of *Salmonella enterica* isolates in genometrakr? From that larger metadata file we would choose the subset we wanted to download and it just seemed like the isolate browser could provide the metadata file I was looking for and was just limited by the file size.

Thank you for your time,

-Enrique

---

**From:** Timme, Ruth <Ruth.Timme@fda.hhs.gov>

**Sent:** Thursday, April 25, 2019 7:47 AM

**To:** Doster, Enrique; GenomeTrakr

**Subject:** Re: How can we download a larger metadata file from the ?

Hi Enrique,

Thanks for reaching out. You can grab the entire metadata file for each species at the FTP site, then parse it for what you need.

Location example for most recent Listeria cluster:

FTP: ftp.ncbi.nlm.nih.gov

Navigate to this directory:

/pathogen/Results/Listeria/PDG000000001.1234/Metadata

There are xml or tsv file options.

Hope this helps!

-Ruth

---

**Ruth E. Timme, PhD**

*Research Microbiologist*

FDA Center for Food Safety and Applied Nutrition

Office of Regulatory Science, Division of Microbiology

College Park, MD 20740

Telework location:

Oakland, CA

FDA mobile: (240) 381-6974

---

**From:** "Doster, Enrique" <Enrique.Doster@colostate.edu>

**Date:** Wednesday, April 24, 2019 at 5:25 PM

**To:** GenomeTrakr <GenomeTrakr@fda.hhs.gov>

**Subject:** How can we download a larger metadata file from the ?

Hello people of GenomeTrakr,

I hope this email finds everyone well. I was hoping you could help me with a question. Our team is attempting to download a big portion of the sequencing data for isolates in the GenomeTrakr database and we are struggling to find the best way to download all the metadata file included with genometrakr. We have tailored our search for the isolate browser here: <https://www.ncbi.nlm.nih.gov/pathogens> but since the metadata file download option is limited to 10k, I'm not able to get all the correct metadata. Is there a way that you all recommend for getting a larger metadata file? Or is there a place to download all of the metadata for us to get the relevant information from?

Thank you in advance for your help,

-Enrique