Methods and tips for extracting data from legume papers.

Order of CSVs for data extraction:

1. “legume\_strain diversity\_meta analysis\_paper information.csv”
2. “legume\_strain diversity\_meta analysis\_plant data.csv”
3. “legume\_strain diversity\_meta analysis\_plant associations\_edited names\_presence absence.csv”
4. “legume\_strain diversity\_meta analysis\_strain sequences.csv”

Reminders:

* In “legume\_strain diversity\_meta analysis\_plant data.csv”, do not fill out columns H – M
* In “legume\_strain diversity\_meta analysis\_plant associations\_edited names\_presence absence.csv” do not fill out columns G-I or M-X
* For “legume\_strain diversity\_meta analysis\_strain sequences.csv” you may need to look up the strain numbers associated with each accession number. You can do this using Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>)
* Use the meta-data for each file in the Data>Meta\_Data folder for reminders on what each column means