**Project Data Explained**

The raw data files used in this project are too large to back up on GitHub. Therefore, the locations of where the data can be found as well as exactly what sequences were used will be described here.

**Soil**

The soil microbiome data was pulled from the Global Atlas of the Dominant Bacteria Found in Soil which can be found at Figshare - https://figshare.com/ under that title. This dataset includes metadata explaining the environmental labels such as what region the biome data is from and the conditions from which the sample were gathered, as well as metadata explaining the dominant bacterial taxa. The dataset contains ten soil samples from Asia and eighty-eight soil samples from North America, the two regions of interest. All samples from Asia were used, since North America had so many associated samples a subset was used.

|  |  |
| --- | --- |
| **North America – Soil Sequence ID’s** | **Asia – Soil Sequence ID’s** |
| 2–24  30-31 | 187-196 |

**Apple/Corn**

Exact SRR’s used from these four datasets (some of the datasets were larger than others so subsets were used in order to keep the comparisons more equal).

|  |  |  |  |
| --- | --- | --- | --- |
| **Apple – Asia** | **Apple – North America** | **Corn – Canada** | **Corn - Asia** |
| [SRR13911931](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR13911931)  [SRR13911926](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR13911926)  [SRR13911925](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR13911925)  [SRR13911924](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR13911924)  [SRR13911919](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR13911919)  [SRR13911918](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR13911918) | [SRR8652343](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652343)  [SRR8652344](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652344)  [SRR8652345](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652345)  [SRR8652346](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652346)  [SRR8652347](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652347)  [SRR8652348](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652348)  [SRR8652349](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652349)  [SRR8652350](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652350)  [SRR8652351](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652351)  [SRR8652352](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652352)  [SRR8652353](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652353)  [SRR8652354](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652354)  [SRR8652355](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652355)  [SRR8652356](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR8652356) | [SRR12525243](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525243)  [SRR12525244](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525244)  [SRR12525245](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525245)  [SRR12525246](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525246)  [SRR12525253](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525253)  [SRR12525254](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525254)  [SRR12525255](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525255)  [SRR12525263](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525263)  [SRR12525264](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525264) [SRR12525265](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525265)  [SRR12525273](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525273)  [SRR12525274](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR12525274) | [SRR18408973](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408973)  [SRR18408971](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408971)  [SRR18408967](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408967)  [SRR18408965](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408965)  [SRR18408955](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408955)  [SRR18408953](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408953)  [SRR18408951](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408951)  [SRR18408949](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408949)  [SRR18408937](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408937)  [SRR18408935](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408935)  [SRR18408933](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408933)  [SRR18408931](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR18408931) |

Sometimes the SRA links don’t work so I have included the SRP’s for manual search if needed.

**Apple – Asia**

Link to SRA Run Selector : <https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP309936&o=acc_s%3Aa>

[SRP309936](https://trace.ncbi.nlm.nih.gov/Traces/?study=SRP309936)

**Apple – North America**

Link to SRA Run Selector : https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP171212&o=acc\_s%3Aa

[SRP171212](https://trace.ncbi.nlm.nih.gov/Traces/sra?study=SRP171212)

**Corn - Asia**

Link to SRA Run Selector : <https://www.ncbi.nlm.nih.gov/Traces/study/?query_key=15&WebEnv=MCID_63ed64c684df6328643a3a58&o=acc_s%3Aa2>

[SRP365154](https://trace.ncbi.nlm.nih.gov/Traces/sra?study=SRP365154)

**Corn – North America**

Link to SRA Run Selector : <https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP278897&o=organism_s%3Aa%3Bacc_s%3Aa>

[SRP387573](https://trace.ncbi.nlm.nih.gov/Traces/?study=SRP387573)