In our example analysis, we investigate the differences between the microbiome of 20 rural and 20 recently urbanized subjects from the Chinese province of Hunan. For more information on this dataset, please review the analysis Fodor Lab published in the Sep 2017 issue of the journal Microbiome: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-017-0338-7

Step 1: Prepare BioLockJ Config File

The BioLockJ project Config chinaKrakenFullDB.properties lists 5 BioModules to run (lines 3-7) + 13 properties:

```
#BioModule biolockj.module.implicit.RegisterNumReads
#BioModule biolockj.module.classifier.wgs.KrakenClassifier
#BioModule biolockj.module.report.taxa.NormalizeTaxaTables
#BioModule biolockj.module.report.r.R_PlotPvalHistograms
#BioModule biolockj.module.report.r.R_PlotOtus
```

In addition to the 5 listed BioModules, 4 additional implicit BioModules will also run:

Mod#	Module	Description
1	ImportMetadata	Always run 1st (for all pipelines)
2	KrakenParser	Always run after KrakenClassifier
3	AddMetadataToOtuTables	Always run just before the 1st R module
4	CalculateStats	Always run as the 1st R module.

Key properties:

Line#	Property	Description
08	cluster.jobHeader	Each script will run on 1 node, 16 cores, and 128GB RAM for up to 30 minutes
10	pipeline.defaultProps	Default config file defines most properties – in this case copperhead.properties
12	input.dirPaths	Directory path containing 40 gzipped whole genome sequencing (WGS) fastq files
18	metadata.filePath	Metadata file path: chinaMetadata.tsv

BioLockJ must associate sequence files in *input.dirPaths* with the correct metadata row. This is done by matching sequence file names to the 1st column in the metadata file. If the Sample ID is not found in your file names, the file names must be updated. Use the following properties to ignore a file prefix or suffix when matching the sample IDs.

- *input.suffixFw*
- *input.suffixRv*
- *input.trimPrefix*
- *input.trimSuffix*