Inputs: parameter file, counts file, metadata file

Parameter file:
The parameter file must include values for the following parameters.

- minSample: samples must be missing from less than X loci
- minloci: loci must be absent from less than Y samples
- umapSeed: RNG seed for UMAP
- epsilon: epsilon value for DBSCAN
- cutHeight: cutoff value for cutting a dendrogram into clusters
- admixedCutoff: clades without a reference and a minimum divergence value above this will be labeled as admixed, null will be interpreted by JSON as None
- filePrefix: prefix for output filenames
- inputCountsFile: name and path to DArT counts file
- inputMetaFile: name and path to the metadata file paired with the countsFile

Example parameter file:
{"minSample":0.2,
"minloci":0.2,
"umapSeed":42,
"epsilon":0.75,
"cutHeight":0.43,
"admixedCutoff":null,
"filePrefix":"riceTutorial",
"inputCountsFile “countsRiceTutorial.csv",
"inputMetaFile":" metadataRiceTutorial.csv"}

Counts file:
Each column is a different samples, and each row is the observed counts for one allele at a given marker.

There must be a column labeled “MarkerName” that contains a label for each marker. Alternate alleles should have the same marker name, and the code assumes that there are exactly two possible alleles for each marker.

The names for the samples should correspond to the short_name column in the metadata file.

Example counts file:

MarkerName	10000	10001	10002	10003
Marker1	0	0	0	0
Marker1	50	30	25	45
Marker2	50	60	10	85
Marker2	0	0	0	0
Marker3	110	220	250	170

Marker3	0	0	0	0
Marker4	0	0	0	0
Marker4	0	0	0	4
Marker5	1	0	0	0
Marker5	1	0	0	0

Metadata file:

The required columns are:

- short_name: sample ID number. This is an integer and must be unique for each row and should be the column label in the counts file.
- reference: the name of the variety if the sample is a reference or None for a field sample. REMOVE indicates that a reference will be removed.

Optional columns:

- inventory: biologic replicate ID number
- reference_original: Sometimes references are submitted with an incorrect variety name. The “reference_original” column should contain whatever variety name was submitted, and any corrections should be made in the “reference” column, which is used to label field samples. This is described in more detail on the reference cleanup page.
- seedSource: A description of where a reference came from. This is one of the possible x-axis labels for the heatmapReferences function, and can be helpful for identifying when references with different genetic fingerprints are different sources.

Example metadata file:

short_name	inventory	reference_original	reference	seedSource
10000	R1	Variety-1	Variety-1a	SourceA
10001	R1	Variety-1	Variety-1a	SourceA
10002	R2	Variety-2	Variety-2	SourceB
10003	R3	None	None	None