

Format of upload spreadsheets

Uploading data in the form of entire spreadsheets requires some pretty finicky formatting and data checking. Mostly, getting a spreadsheet to 'check' for upload requires an iterative approach where you fix problems the upload checker identifies

Format of spreadsheet file

- The spreadsheet has to be in the form of a comma separated value (*.csv) file.
- The spreadsheet has to have specific column names present in the first row (though order is not important).

Required columns

There are several columns that need to be included in each spreadsheet that also contains phenotypic data. These correspond broadly to the experimental treatments, locations of experiment, identity of the plant (and its parent) and anecdotal comments about that particular plant.

These columns need data. If they are left blank, it makes it hard to perform all the cool analyses that these data will allow. Please try very hard to enter accurate, reasonable values for each plant.

Here is a list of columns with some descriptions of their contents

accession The name of the line. Most of the unPAK line ids are in the form of 'SALK_nnnnnnC' or 'CS70000'

expt.id The name of the experiment. Usually these are agreed upon across institutions or within a CURE or set of CUREs. See [the list of experiments](#)

institution The name of the institution. See [the list of institutions](#)

facility The ID of the growth chamber or greenhouse See [the list of facilities](#)

expt.plantnum the unique number given to this plant during the experiment

flat the flat it was grown in

row the row it was grown in

column the column ...

date date that the phenotype was finalized

investigator

parent.id The id for the parent of the plant. These are constructed from the institution, experiment name and the plantnum of the parent during that experiment.

Phenotypes

At least one column specifying a numeric phenotype is required. If there are multiple columns with numeric phenotypic data, they are all used.

The name of the column should correspond to existing phenotypes in the database See [the list of phenotypes currently in the database](#) .

If the phenotype does not exist in the db currently, then an error is generated.

This will happen when you have a phenotype new to unpak. In this case email `unpak.database@gmail.com` with the information about your new phenotype. This information must include all of the information for your new phenotype that corresponds to current entries in the [phenotype table](#) .

Please make sure that your new phenotype really is different than an existing one. One great strength of unPAK is the fact that we can compare phenotypes across experiments. If your phenotype has a different name, but measures the same thing as one already in the db, we will not be able to take full advantage of your hard work.

Optional columns

These columns are not required, but information in them will be recorded.

comment The is a comment that corresponds to this plant. It might describe some specific problem with the plant or its phenotypes.

Questions or requests for changes to db

email `unpak.database@gmail.com`