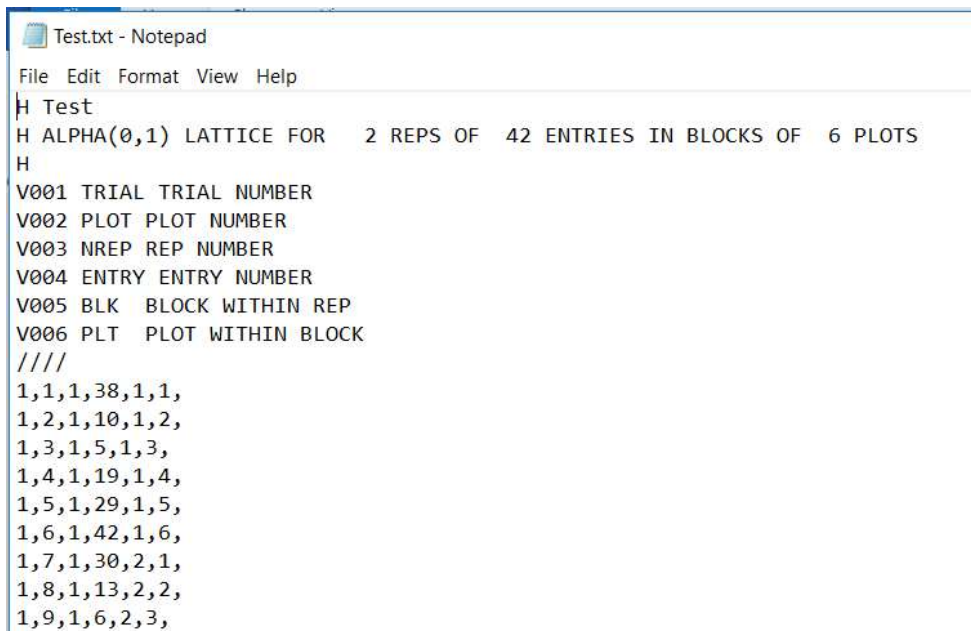


Tutorial on Importing an existing design for large trials

Importing a basic design

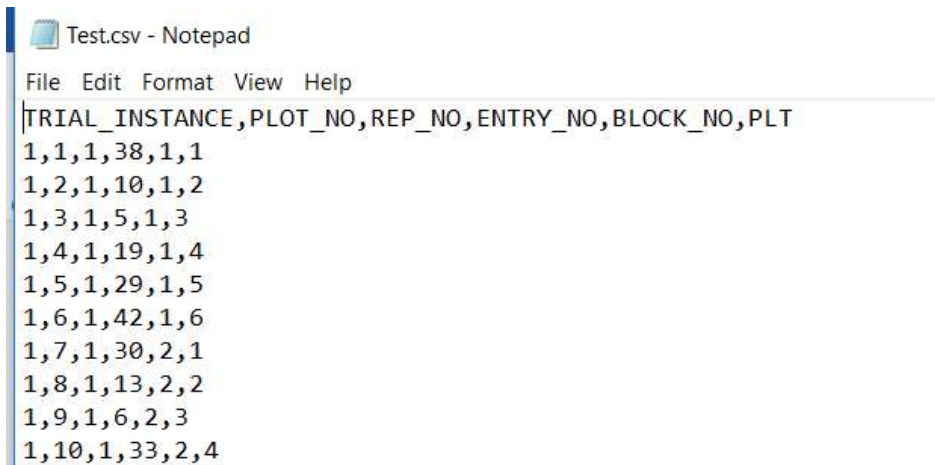
The current version of BMS (V5) has a limitation on the number of environments (trial instances) that can be generated with the trial design engine. The current limit is about 25 instances. This will be increased in a future release. In the mean time it is possible to create larger trials using the “Import an Existing Design” feature of the trial manager.

1. Generate a layout file with your favorite design software. For example with CropStat you can generate an Alpha design with two reps of 42 entries for 80 locations. If you open the .sys layout file in the CropStat data editor, you can export the file as .txt.



```
Test.txt - Notepad
File Edit Format View Help
H Test
H ALPHA(0,1) LATTICE FOR 2 REPS OF 42 ENTRIES IN BLOCKS OF 6 PLOTS
H
V001 TRIAL TRIAL NUMBER
V002 PLOT PLOT NUMBER
V003 NREP REP NUMBER
V004 ENTRY ENTRY NUMBER
V005 BLK BLOCK WITHIN REP
V006 PLT PLOT WITHIN BLOCK
////
1,1,1,38,1,1,
1,2,1,10,1,2,
1,3,1,5,1,3,
1,4,1,19,1,4,
1,5,1,29,1,5,
1,6,1,42,1,6,
1,7,1,30,2,1,
1,8,1,13,2,2,
1,9,1,6,2,3,
```

2. To import this to BMS you need to replace all the header text with the line shown in the figure below and save the file with extension .csv. The column headings must match **EXACTLY** the variable names for the design variables in the ontology.



```
Test.csv - Notepad
File Edit Format View Help
TRIAL_INSTANCE,PLOT_NO,REP_NO,ENTRY_NO,BLOCK_NO,PLT
1,1,1,38,1,1
1,2,1,10,1,2
1,3,1,5,1,3
1,4,1,19,1,4
1,5,1,29,1,5
1,6,1,42,1,6
1,7,1,30,2,1
1,8,1,13,2,2
1,9,1,6,2,3
1,10,1,33,2,4
```

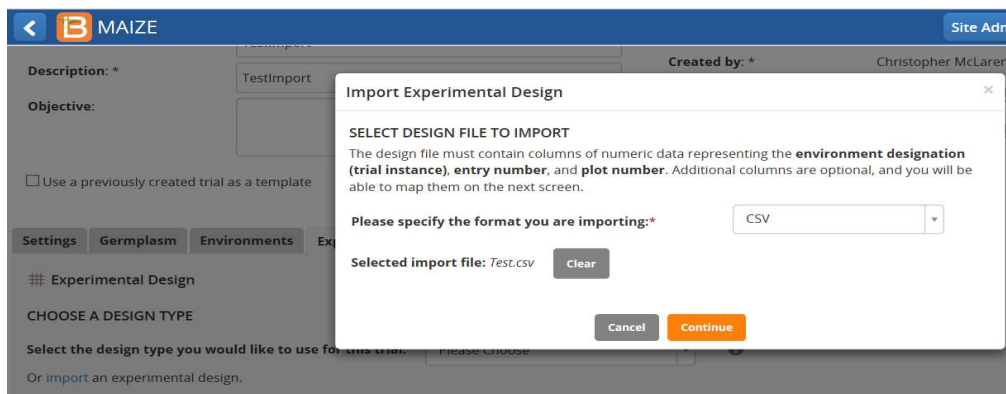
- This csv file can be opened in excel. Delete the last column (PLT) since this is not used by BMS for an Alpha Lattice design

	A	B	C	D	E	F	G
1	TRIAL_INSTANCE	PLOT_NO	REP_NO	ENTRY_NO	BLOCK_NO	PLT	
2	1	1	1	38	1	1	
3	1	2	1	10	1	2	
4	1	3	1	5	1	3	
5	1	4	1	19	1	4	
6	1	5	1	29	1	5	
7	1	6	1	42	1	6	
8	1	7	1	30	2	1	
9	1	8	1	13	2	2	
10	1	9	1	6	2	3	
11	1	10	1	33	2	4	
12	1	11	1	41	2	5	
13	1	12	1	9	2	6	
14	1	13	1	37	3	1	

Then save the lay-out file.

	A	B	C	D	E	F
1	TRIAL_INSTANCE	PLOT_NO	REP_NO	ENTRY_NO	BLOCK_NO	
2	1	1	1	38	1	
3	1	2	1	10	1	
4	1	3	1	5	1	
5	1	4	1	19	1	
6	1	5	1	29	1	
7	1	6	1	42	1	
8	1	7	1	30	2	
9	1	8	1	13	2	
10	1	9	1	6	2	
11	1	10	1	33	2	
12	1	11	1	41	2	

- In BMS go to Trial Manager and Create a New Trial. Either use a template trial or design the trial from scratch. On the Germplasm tab select the trial germplasm (with 42 entries for this example), on the Environments tab enter 80 environments. Then on the design tab select 'import' and select the layout file in csv format as above.



5. Check the mapping and click Next:

The screenshot displays a trial configuration interface with three sections of factors:

- Environmental Factors** (1 factor):
 - TRIAL_INSTANCE → TRIAL_INSTANCE (Required)**: Property: Trial instance, Scale: Number, Method: Enumerated. Includes a Re-map button.
- Design Factors** (3 factors):
 - PLOT_NO → PLOT_NO (Required)**: Property: Field plot, Scale: Number, Method: Enumerated. Includes a Re-map button.
 - REP_NO → REP_NO**: Property: Replication factor, Scale: Number, Method: Enumerated. Includes a Re-map button.
 - BLOCK_NO → BLOCK_NO**: Property: Blocking factor, Scale: Number, Method: Enumerated. Includes a Re-map button.
- Germplasm Factors** (1 factor):
 - ENTRY_NO → ENTRY_NO (Required)**: Property: Germplasm entry, Scale: Number, Method: Enumerated. Includes a Re-map button.

At the bottom, there are "Cancel" and "Next" buttons.

6. Review the details of the imported measurements and click Finish.
7. After the first Success message (which tells you that the table was imported) you **MUST** click Save to save the imported trial.

The screenshot shows the "MANAGE TRIALS" interface with a "Create Trial" form and a sidebar:

- MANAGE TRIALS** header with a help icon.
- Create Trial** section with a "Save" button.
- BASIC DETAILS** section with a note: ** indicates a mandatory field*.
- Trial name: *** field containing "TestImport".
- Description: *** field containing "TestImport".
- Objective:** empty text area.
- Trials** sidebar showing a folder structure: Trials > 2017 MM > CGM (highlighted) > JP > Templates.

This saving takes some time and there is no blue wheel to help so be patient!

8. You can now return to manage trials, open this trial, add environmental conditions such as LOCATION_NAME or Crop_season_Code on the environments tab. You can also add the traits you want to measure on the measurements tab etc.

Getting more out of Importing a design

In fact the import design function is very powerful and you can do a lot more than the basic import shown above. For example you can import Environmental details and conditions, traits and even measurements.

Importing Environment Conditions

Adding the environmental details like LOCATION_NAME for 80 locations can be quite a challenge. However if you have a list of the locations in excel you can import them with the design.

1. Load the layout file into an xlsx file. Name the sheet Layout. Add a second sheet called Locations and add a list of locations as in the screenshot below:

	A	B
1	TRIAL_INSTANCE	LOCATION_NAME
2	1	Kitale Research Station
3	2	CHIMOIO
4	3	GART
5	4	PETAUKE
6	5	SERENJE
7	6	MUMBWA
8	7	KAFUE
9	8	Chibero
10	9	BOTSWANA (RS)
11	10	CHINGOLA
12	11	KILOLO
13	12	MBEYA
14	13	IRINGA
15	14	MERU
16	15	MKUSHI
17	16	KABWE (ZARI)
18	17	CHINHOYI UNIVERSITY
19	18	ZAKA

There are 80 trial instances in this list!

Note: BMS will search in the database for the locations to get their LocIDs. **The names MUST be exactly as they are in the database** (Manage Locations), and BMS will take the first one to match the name – so you should **always make unique trial location names**. If location names are not found in the database then no location name is stored for that environment. No error is given.

2. Next on the Layout sheet insert columns for LOCATION_NAME and Crop_season_Code. (The names of the variables must match exactly the names in your ontology). You can fill the LOCATION_NAME column using the VLOOKUP formula that you can see in the screenshot below:

	A	B	C	D	E	F	G
1	TRIAL_INSTANCE	LOCATION_NAME	Crop_season_Code	PLOT_NO	REP_NO	ENTRY_NO	BLOCK_NO
2	1	Kitale Research Station	17A	1	1	38	1
3	1	Kitale Research Station	17A	2	1	10	1
4	1	Kitale Research Station	17A	3	1	5	1
5	1	Kitale Research Station	17A	4	1	19	1
6	1	Kitale Research Station	17A	5	1	29	1
7	1	Kitale Research Station	17A	6	1	42	1
8	1	Kitale Research Station	17A	7	1	30	2
9	1	Kitale Research Station	17A	8	1	13	2
10	1	Kitale Research Station	17A	9	1	6	2

- Fill the Crop_season_Code with valid values exactly as in the ontology. (The same LOCATION_NAME and Crop_season_Code for every plot in each trial instance).
- Save the Layout sheet as a csv file from excel.
- Make sure the variables LOCATION_NAME and Crop_season_Code are added to the Environment details section of the Environments tab. (If you don't you will get an unhelpful error when you try to import the design "cannot persist germplasm list")
- Set up a trial and import the design just as in the previous section and the LOCATION_NAME and Crop_season_Code values will be imported into the Environments dataset.

Importing Trait Names and Values

- If you add column headings to the layout file which correspond EXACTLY to the names of trait variables in the ontology they will be imported to the measurements tab.

	A	B	C	D	E	F	G	H	I	J
1	TRIAL_INSTANCE	LOCATION_NAME	Crop_season_Code	PLOT_NO	REP_NO	ENTRY_NO	BLOCK_NO	GY_Adj_tha	AflavER_1_5	
2	1	Kitale Research Station	1	1	1	38	1	10.92	3	
3	1	Kitale Research Station	1	2	1	10	1	6.08	4	
4	1	Kitale Research Station	1	3	1	5	1	12.29	3	
5	1	Kitale Research Station	1	4	1	19	1	9.79	4	
6	1	Kitale Research Station	1	5	1	29	1	5.56	4	
7	1	Kitale Research Station	1	6	1	42	1	11.39	3	
8	1	Kitale Research Station	1	7	1	30	2	8.17	4	
9	1	Kitale Research Station	1	8	1	13	2	11.93	3	
10	1	Kitale Research Station	1	9	1	6	2	16.05	3	

- If also the columns under the trait names are filled with VALID values they will be imported to the measurements tab.

Generating Random Data for Traits

Sometimes for teaching or testing BMS (analysis or Breeders Queries) it is useful to have a lot of practice data. However since the randomizations and layouts of trials in BMS are all unique it is very difficult to generate data which matches the design and shows genetic effects and correlations. However excel can be used to generate some interesting data examples.

- In the xlsx file used to create the layout described above add two more sheets: Germplasm and Parameters.
- In the Germplasm sheet enter the columns ENTRY_NO and GEEF. Fir the first with integers 1,2, ...42 and the second with the formula: =NORM.INV(RAND(),0,2). This formula generated a random genotype effect for the corresponding entry from a Normal Distribution with mean 0 and standard deviation 2 (Variance 4). Copy the formula down the list of 42 entries.

	A	B	C	D	E	F
1	ENTRY_NO	GEEF				
2	1	1.677639				
3	2	2.763327				
4	3	0.084576				
5	4	0.683434				
6	5	-0.00687				
7	6	-0.37886				
8	7	0.298883				
9	8	1.140418				
10	9	-0.39254				
11	10	-0.59024				

- In the Parameters sheet enter the column headings BLOCK_NO and BKEF. Fill the BLOCK_NO column with integers 1, ...7 since we have 7 blocks per rep. Fill the BKEF column with Normal random numbers with mean 0 and variance 4 with the formula shown in the screenshot below.

	A	B	C	D	E	F	G
1	BLOCK	BKEF					
2	1	0.233784					
3	2	2.769657					
4	3	-2.20459					
5	4	1.739696					
6	5	-3.8759					
7	6	0.591603					
8	7	0.433718					
9							

You can vary the SD in this formula to give larger or smaller block effects. The same effect is used for all reps and all locations which is not very realistic but it is not too bad.

- Now on the Layout Sheet you can fill the Yield column (GY_Adj_tha) with random numbers with the formula: $=\text{MAX}(0, \text{TRUNC}(10 + \text{A2}/10 + 2 * (2 * \text{TRUNC}(\text{A2}/2, 0) - \text{A2} + 0.5) * \text{VLOOKUP}(\text{F2}, \text{Germplasm}!\$A\$2:\$B\$43, 2, \text{FALSE}) + \text{E2} + \text{VLOOKUP}(\text{G2}, \text{Parameters}!\$A\$2:\$B\$8, 2, \text{FALSE}) + \text{NORM.INV}(\text{RAND}(), 0, 3), 2))$

- The MAX clause simply puts 0 in cases where the random number comes out negative (roughly 1 in 5000 with my parameter settings).
- $10 + \text{A2}/10 +$ this sets the grand mean to 10 + Trial instance no/2. So 10.1 for trial 1 and 18 for trial 80.
- $2 * (2 * \text{TRUNC}(\text{A2}/2, 0) - \text{A2} + 0.5)$ this factor is -1 for trials with odd TRIAL_INSTANCE number and +1 for even.
- $\text{VLOOKUP}(\text{F2}, \text{Germplasm}!\$A\$2:\$B\$43, 2, \text{FALSE})$ This looks up the genotype effect from the Germplasm sheet according to the ENTRY_NO (in column F). (The \$ characters are necessary to fix the search range on the Germplasm sheet so that it remains the same when you copy the formula down the Yield column.
- The product of the last two factors ensures that the genetic correlation between odd and even numbered trial instances is -1.
- +E2+ is a complete block effect (1 or 2).
- $\text{VLOOKUP}(\text{G2}, \text{Parameters}!\$A\$2:\$B\$8, 2, \text{FALSE})$ This factor looks up a random sub-block effect from the parameters sheet.
- $\text{NORM.INV}(\text{RAND}(), 0, 3), 2)$ The last factor is the error variance Normal with mean 0 and variance 4. Again you can change the standard deviation (2 in the example) to give different repeatability values. The example settings give average repeatability of .33.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	TRIAL_INSTANCE	LOCATION_NAME	Crop_season_Code	PLOT_NO	REP_NO	ENTRY_NO	BLOCK_NO	GY_Adj_tha	AflavER_1_5								
2	1	Kitale Research Station	17A	1	1	38	1	10.21	3								
3	1	Kitale Research Station	17A	2	1	10	1	5.59	4								
4	1	Kitale Research Station	17A	3	1	5	1	13.29	3								
5	1	Kitale Research Station	17A	4	1	19	1	10.19	3								
6	1	Kitale Research Station	17A	5	1	29	1	5.76	4								
7	1	Kitale Research Station	17A	6	1	42	1	9.65	4								
8	1	Kitale Research Station	17A	7	1	30	2	14.09	3								
9	1	Kitale Research Station	17A	8	1	13	2	13.24	3								

- You can fill the AflavER_1_5 column with values between 1 and 5 (as required) with the formula:

$$=6-\text{MIN}(5,\text{TRUNC}(H2/5,0))+1$$

Which divides the range of GY_Adj_tha into 5 parts and sets the plots with the highest yields the lowest Aspergillus flavus ear rot scores and those with the lowest yields the higher scores.

- You need to copy the formula for GY_Adj_tha and for AflavER_1_5 down the complete columns for each trait to generate values for every plot.

	A	B	C	D	E	F	G	H	I
1	TRIAL_INSTANCE	LOCATION_NAME	Crop_season_Code	PLOT_NO	REP_NO	ENTRY_NO	BLOCK_NO	GY_Adj_tha	AflavER_1_5
2	1	Kitale Research Station	17A	1	1	38	1	7.82	4
3	1	Kitale Research Station	17A	2	1	10	1	4.36	5
4	1	Kitale Research Station	17A	3	1	5	1	12.33	3
5	1	Kitale Research Station	17A	4	1	19	1	9.18	4
6	1	Kitale Research Station	17A	5	1	29	1	14.05	3
7	1	Kitale Research Station	17A	6	1	42	1	7.72	4
8	1	Kitale Research Station	17A	7	1	30	2	9.64	4

- Now you can save the Layout sheet of this excel file as a csv file (which converts the formulae to values) and you can import that file as described in the sections above.