Using Cvt2Mae to Convert GenePix Array Data for MAExplorer

http://www.lecb.ncifcrf.gov/Cvt2Mae

Peter F. Lemkin(1), Greg Thornwall (2), Bob Stephens(3)

(1) LECB/NCI/FCRDC, (2) SAIC/FCRDC, (3) ABCC/FCRDC

DRAFT - Revised: 01-28-2002
Cvt2Mae version 0.60

Accessing Arrays with MAExplorer

- MAExplorer works with any arrays using the schema (see Appendix C of MAExplorer Reference Manual for details)
- All data files are tab-delimited text files
- Databases could be constructed with tools like Excel for editing user data into the schema format
- The Cvt2Mae array data converter "Wizard" tool converts non-standard <User-defined> academic or commercial data to MAExplorer format
- Affymetrix, Incyte, **GenePix**, Scanalyze, and other array data formats may be converted using predefined "Array Layouts"

S.1 MAExplorer Data Schema

- MAExplorer works with any array data using our data schema

- The schema is described in detail in MAExplorer Reference Manual Appendix C.

- Data Schema: tab-delimited experiment data files:
  1. GIPO (Gene In Plate Order or "print" file)
  2. List of hybridized samples in database
  3. Configuration data describing the array and conventions
  4. Separate spot quantification data files

- The Cvt2Mae "wizard" tool converts user array data to this schema

S.1.1 MAExplorer GIPO or Print File

- GIPO file maps a spot on the array to a particular gene

- Contains:
  1. location or grid-geometry
  2. one or more genomic identifiers (e.g., Clone ID, GenBank ID, LocusID, etc.)
  3. gene description as Gene Name (or other description)
  4. Optional: global spot quality (QualCheck)
  5. optional: plate coordinates for clones

S.1.2 MAExplorer Samples Database File

- List of hybridized samples file SamplesDB.txt file contains:

1. full sample description

2. base file name of quantification file (without .quant file extension)

3. optional sample ID number

4. other data you wish to carry with the samples (used in array reports)

S.1.3 MAExplorer Configuration Database File

- Configuration data file MaeConfig.txt describes particular type of array and hybridization labeling you are using. This includes:

- grid-geometry - # of replicate fields, grids, rows/grid, columns/grid

- spot hybridization labeling - intensity or ratio data, dye names

- various presentation options - use pseudo-array or actual (x,y) coordinates, etc.

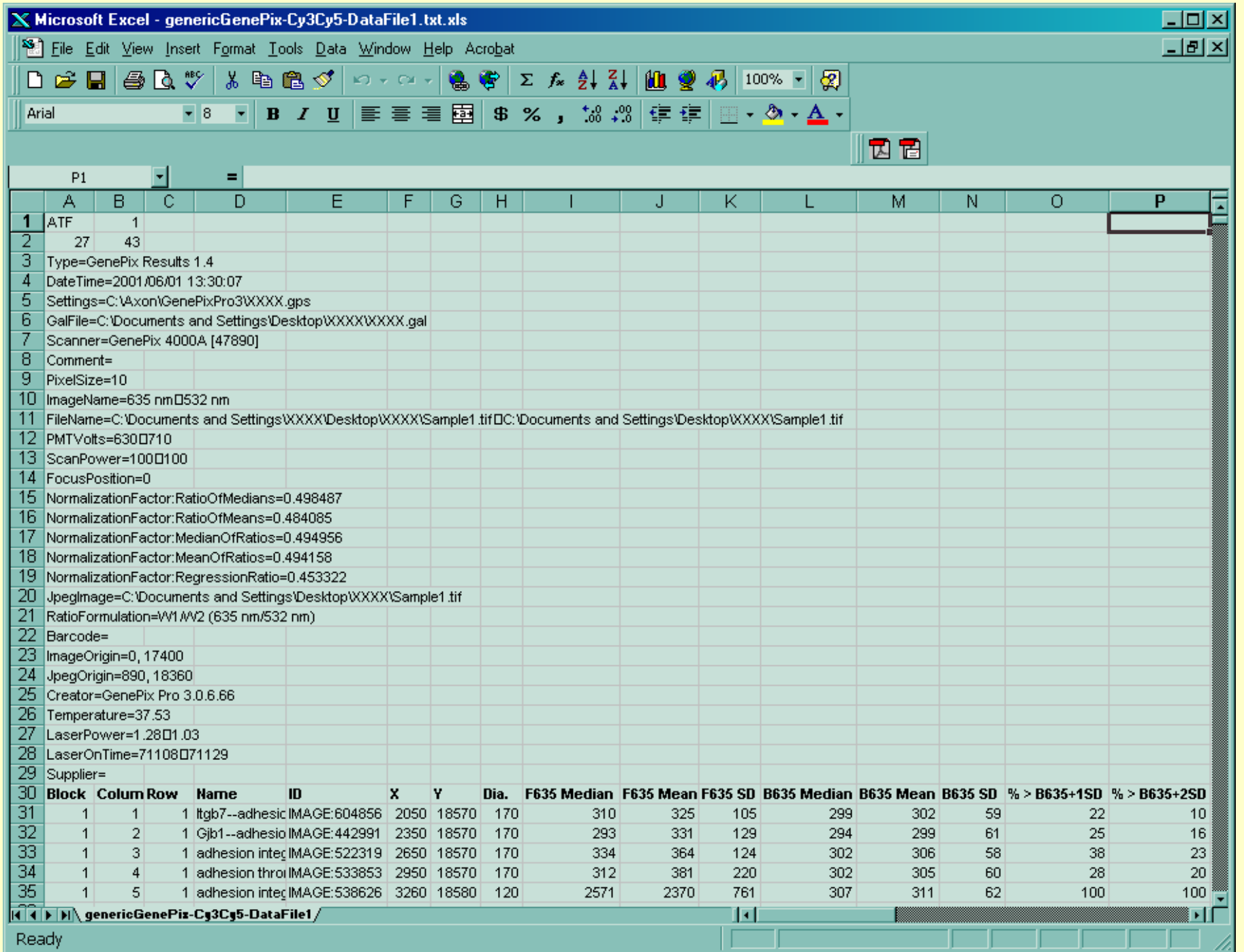
S.1.4 MAExplorer Spot Quantification Files

- Separate spot quantification data (with .quant file extension) files are used for each hybridized sample
- ^{33}P or biotin labeled samples are specified as one hybridization intensity information per file
- Fluorescent Cy3/Cy5-dye labeled samples are specified as two channels of hybridization intensity information per file
- Intensity background data is optional
- Spot quality (QualCheck) data is optional
- Grid-coordinates are specified the same as for GIPO file

S.2 Assumptions About User Data - Array Layout

- User data is tab-delimited ASCII text files (could generate with Excel)
- If the array geometry (#fields, grids, rows/grid, columns/grid) is known, that geometry may be used in MAExplorer
- Otherwise, a pseudo-array geometry is generated for visual use in MAExplorer from the total # of spots in the user data
- An Array Layout describes the user data. It may be edited and saved for subsequent use in converting other array data files of the same type
- The <User-defined> array layout gives users complete flexibility in describing the array

S.3 Example of tab-delimited GenePix Data



Microsoft Excel - genericGenePix-Cy3Cy5-DataFile1.txt.xls

File Edit View Insert Format Tools Data Window Help Acrobat

Arial 8

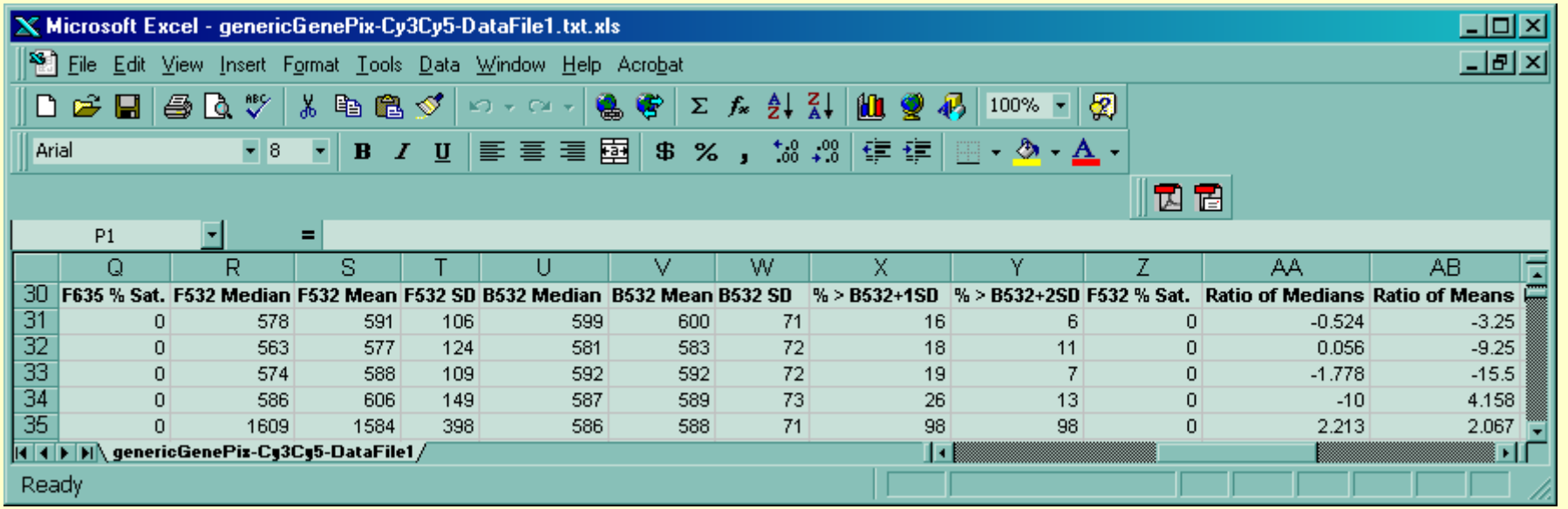
P1 =

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ATF	1														
2	27	43														
3	Type=GenePix Results 1.4															
4	DateTime=2001/06/01 13:30:07															
5	Settings=C:\Axon\GenePixPro3\XXXX.gps															
6	GalFile=C:\Documents and Settings\Desktop\XXXX\XXXX.gal															
7	Scanner=GenePix 4000A [47890]															
8	Comment=															
9	PixelSize=10															
10	ImageName=635 nm□532 nm															
11	FileName=C:\Documents and Settings\XXXX\Desktop\XXXX\Sample1.tif□C:\Documents and Settings\Desktop\XXXX\Sample1.tif															
12	PMTVolts=630□710															
13	ScanPower=100□100															
14	FocusPosition=0															
15	NormalizationFactor:RatioOfMedians=0.498487															
16	NormalizationFactor:RatioOfMeans=0.484085															
17	NormalizationFactor:MedianOfRatios=0.494956															
18	NormalizationFactor:MeanOfRatios=0.494158															
19	NormalizationFactor:RegressionRatio=0.453322															
20	JpegImage=C:\Documents and Settings\Desktop\XXXX\Sample1.tif															
21	RatioFormulation=W1/W2 (635 nm/532 nm)															
22	Barcode=															
23	ImageOrigin=0, 17400															
24	JpegOrigin=890, 18360															
25	Creator=GenePix Pro 3.0.6.66															
26	Temperature=37.53															
27	LaserPower=1.28□1.03															
28	LaserOnTime=71108□71129															
29	Supplier=															
30	**Block**	**Colum**	**Row**	**Name**	**ID**	**X**	**Y**	**Dia.**	**F635 Median**	**F635 Mean**	**F635 SD**	**B635 Median**	**B635 Mean**	**B635 SD**	**% > B635+1SD**	**% > B635+2SD**
31	1	1	1	Itgb7--adhesic	IMAGE:604856	2050	18570	170	310	325	105	299	302	59	22	10
32	1	2	1	Gjb1--adhesio	IMAGE:442991	2350	18570	170	293	331	129	294	299	61	25	16
33	1	3	1	adhesion intec	IMAGE:522319	2650	18570	170	334	364	124	302	306	58	38	23
34	1	4	1	adhesion throi	IMAGE:533853	2950	18570	170	312	381	220	302	305	60	28	20
35	1	5	1	adhesion intec	IMAGE:538626	3260	18580	120	2571	2370	761	307	311	62	100	100

genericGenePix-Cy3Cy5-DataFile1

Ready

S.3.1 Example of tab-delimited GenePix Data



Microsoft Excel - genericGenePix-Cy3Cy5-DataFile1.txt.xls

File Edit View Insert Format Tools Data Window Help Acrobat

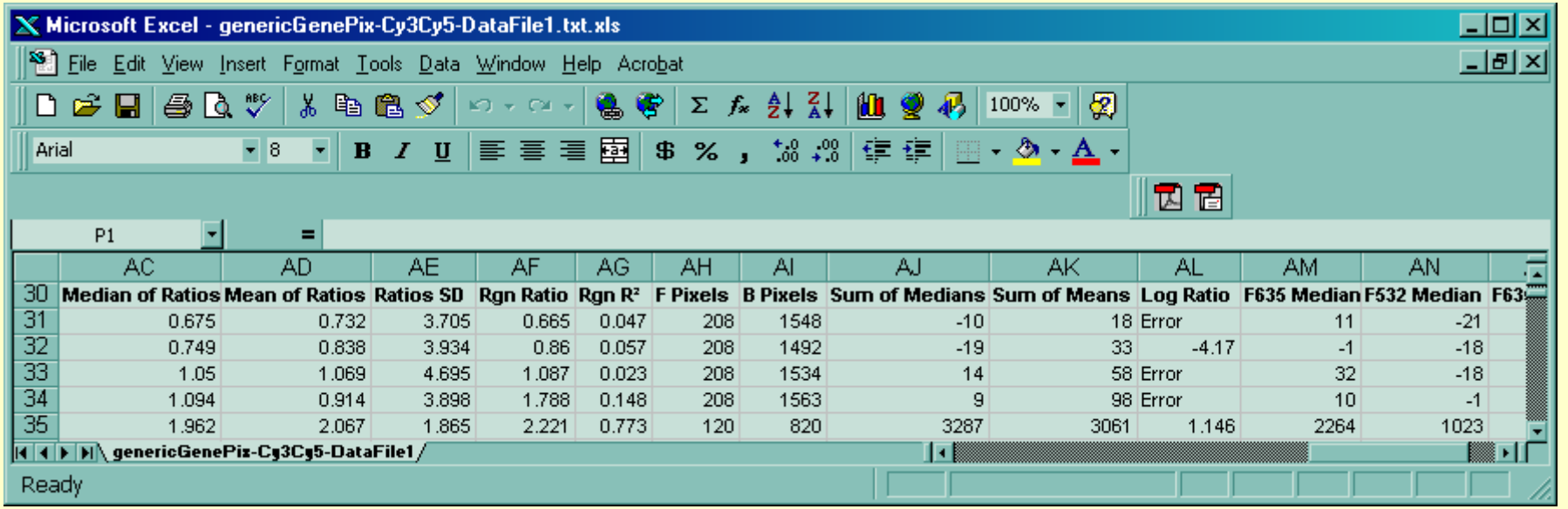
Arial 8 100%

P1 =

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
30	F635 % Sat.	F532 Median	F532 Mean	F532 SD	B532 Median	B532 Mean	B532 SD	% > B532+1SD	% > B532+2SD	F532 % Sat.	Ratio of Medians	Ratio of Means
31	0	578	591	106	599	600	71	16	6	0	-0.524	-3.25
32	0	563	577	124	581	583	72	18	11	0	0.056	-9.25
33	0	574	588	109	592	592	72	19	7	0	-1.778	-15.5
34	0	586	606	149	587	589	73	26	13	0	-10	4.158
35	0	1609	1584	398	586	588	71	98	98	0	2.213	2.067

genericGenePix-Cy3Cy5-DataFile1

Ready



Microsoft Excel - genericGenePix-Cy3Cy5-DataFile1.txt.xls

File Edit View Insert Format Tools Data Window Help Acrobat

Arial 8 100%

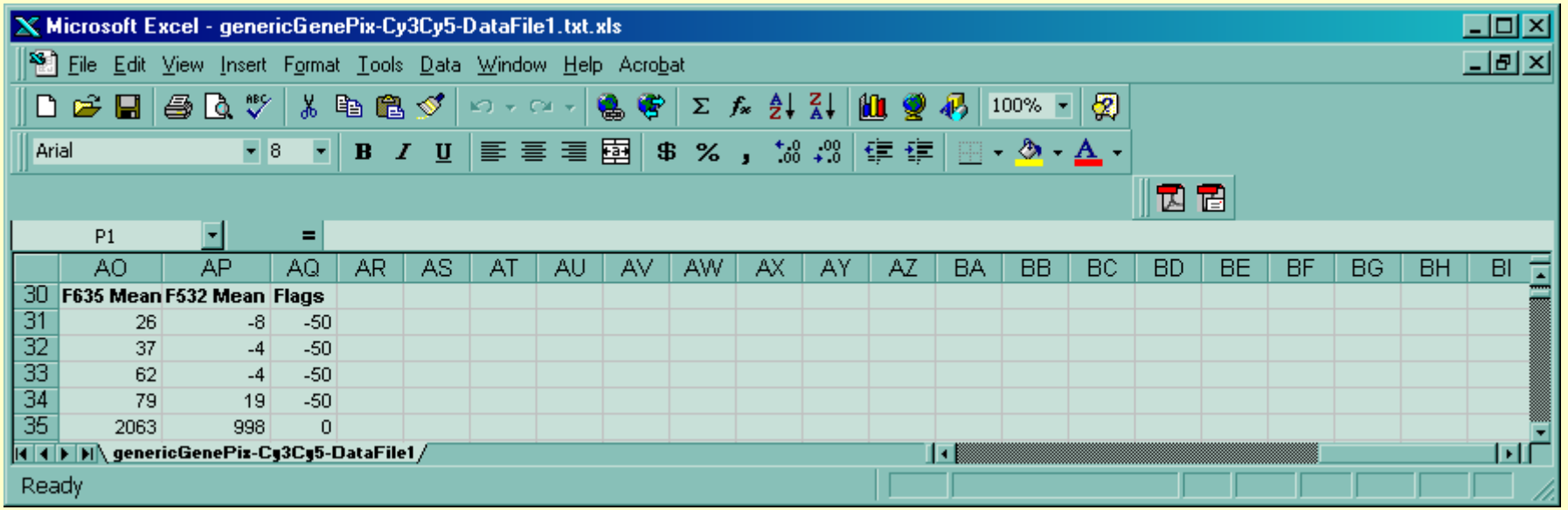
P1 =

	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
30	Median of Ratios	Mean of Ratios	Ratios SD	Rgn Ratio	Rgn R²	F Pixels	B Pixels	Sum of Medians	Sum of Means	Log Ratio	F635 Median	F532 Median
31	0.675	0.732	3.705	0.665	0.047	208	1548	-10	18	Error	11	-21
32	0.749	0.838	3.934	0.86	0.057	208	1492	-19	33	-4.17	-1	-18
33	1.05	1.069	4.695	1.087	0.023	208	1534	14	58	Error	32	-18
34	1.094	0.914	3.898	1.788	0.148	208	1563	9	98	Error	10	-1
35	1.962	2.067	1.865	2.221	0.773	120	820	3287	3061	1.146	2264	1023

genericGenePix-Cy3Cy5-DataFile1

Ready

S.3.2 Example of tab-delimited GenePix Data



	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI
30	**F635 Mean**	**F532 Mean**	**Flags**																		
31	26	-8	-50																		
32	37	-4	-50																		
33	62	-4	-50																		
34	79	19	-50																		
35	2063	998	0																		

I. Procedure: Convert Data for Array Layouts

1. Select the Chip Set array layout (**GenePixPro3 - generic**) if in list, otherwise pick <User-defined>)

2. Select 1 or more input files using the "Browse filename" .

3. You may edit or change various array layout parameters at this time
   3.1 you may edit the array layout with "Edit Layout"
   3.2 you may "Assign GIPO fields" in user data file
   3.3 you may "Assign Quantification fields" in user data file
   3.4 if you changed any array layout parameters, you may save it with "Save Layout"

4. Select the project output directory (i.e., folder) to save generated files

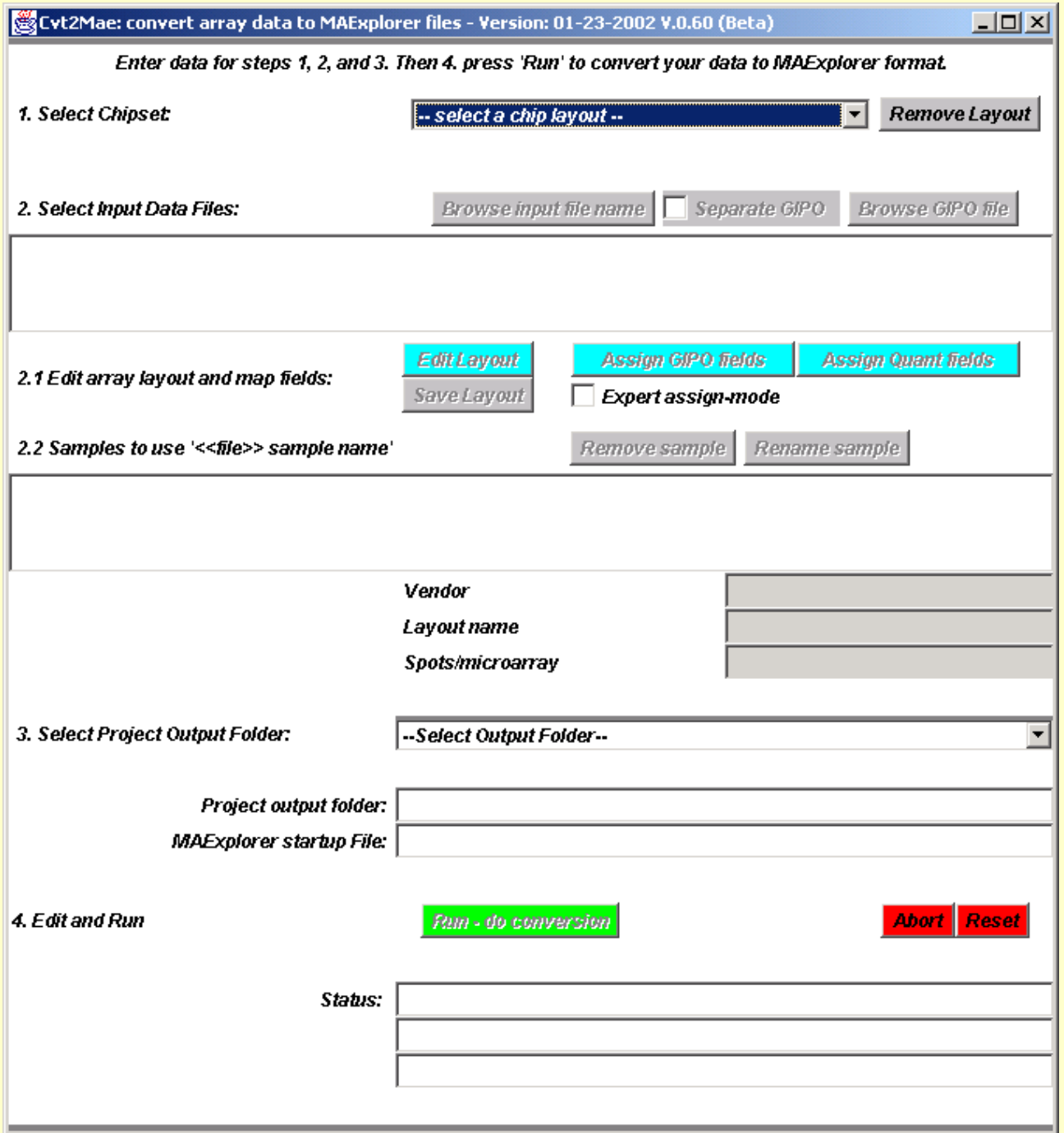
I. Procedure: continued...

5. Press “Run” to convert the data

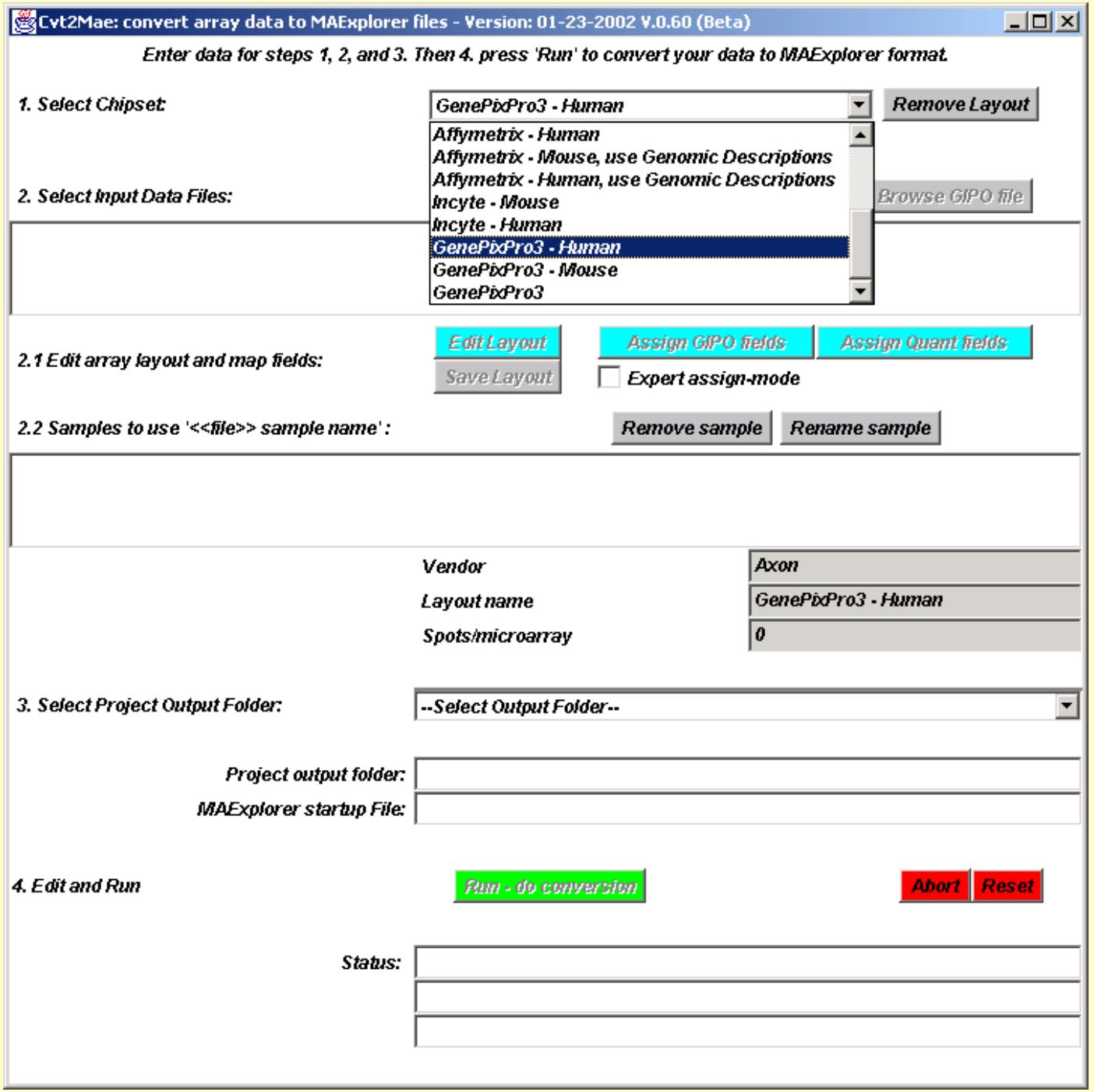
6. Press “Done” when it is finished.

7. Go to the project directory and then to the MAE sub-directory, click on the Start.mae file to start MAExplorer on the new data

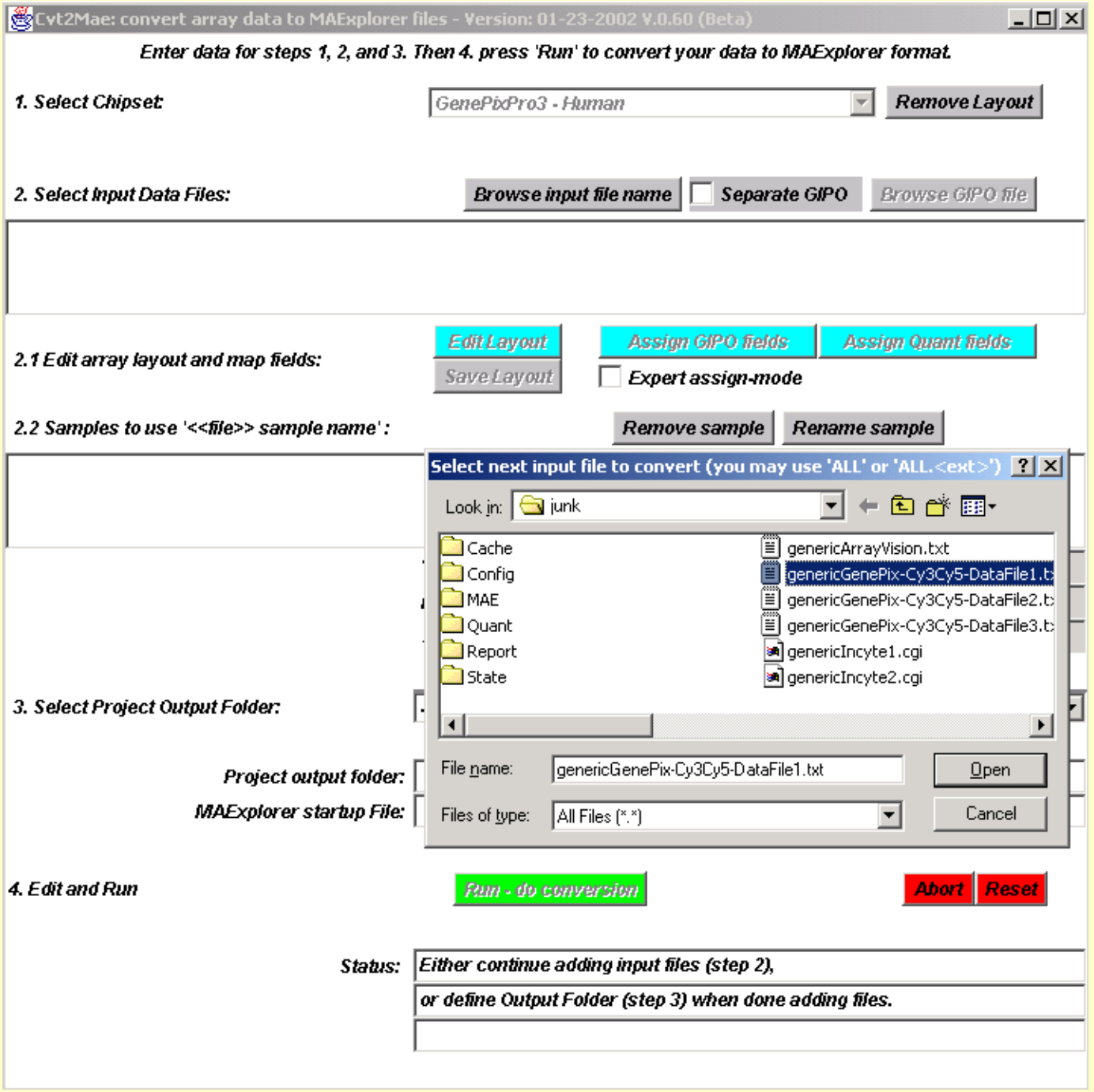
1. Initial State of Cvt2Mae Program



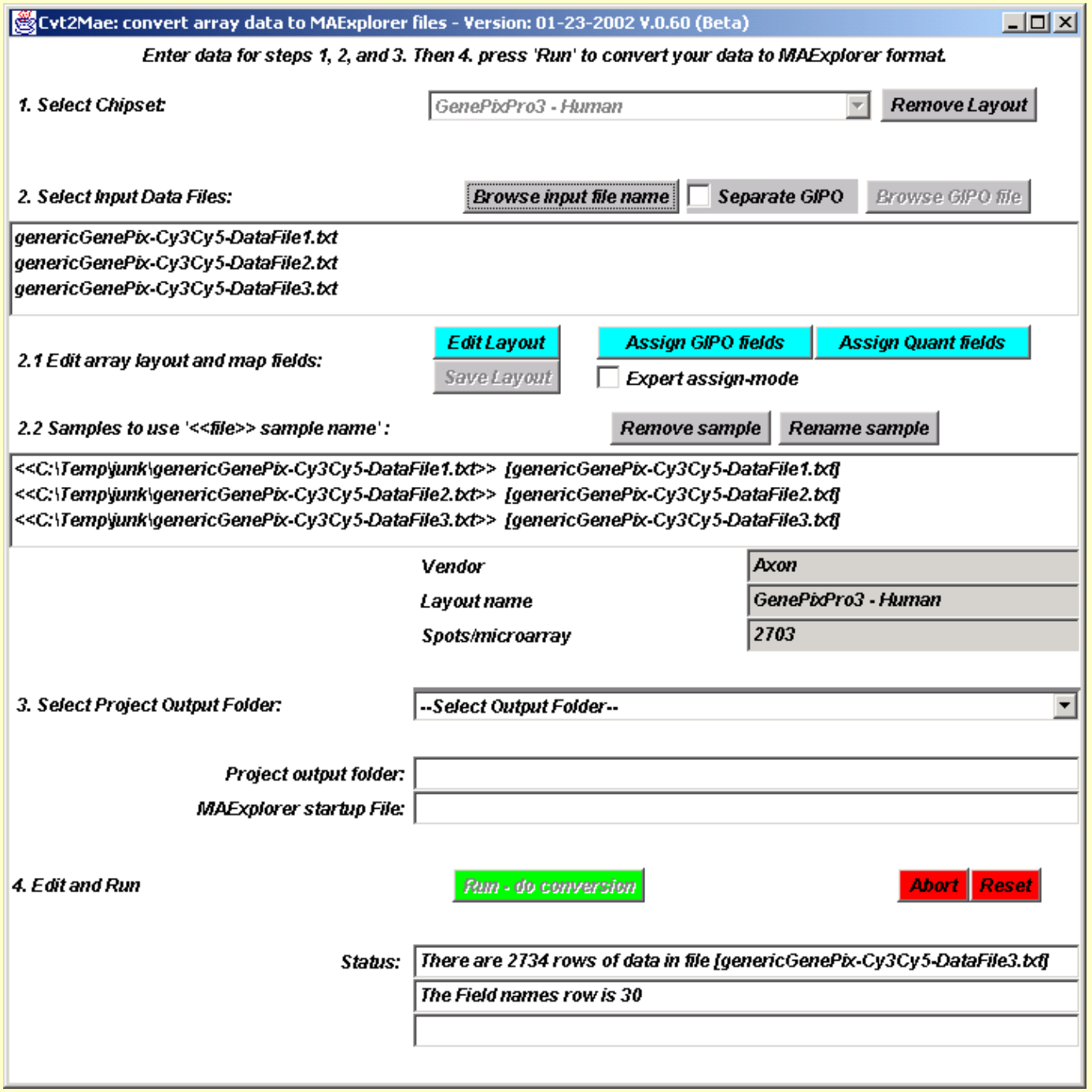
2. Selecting GenePix Chipset Array-Layout



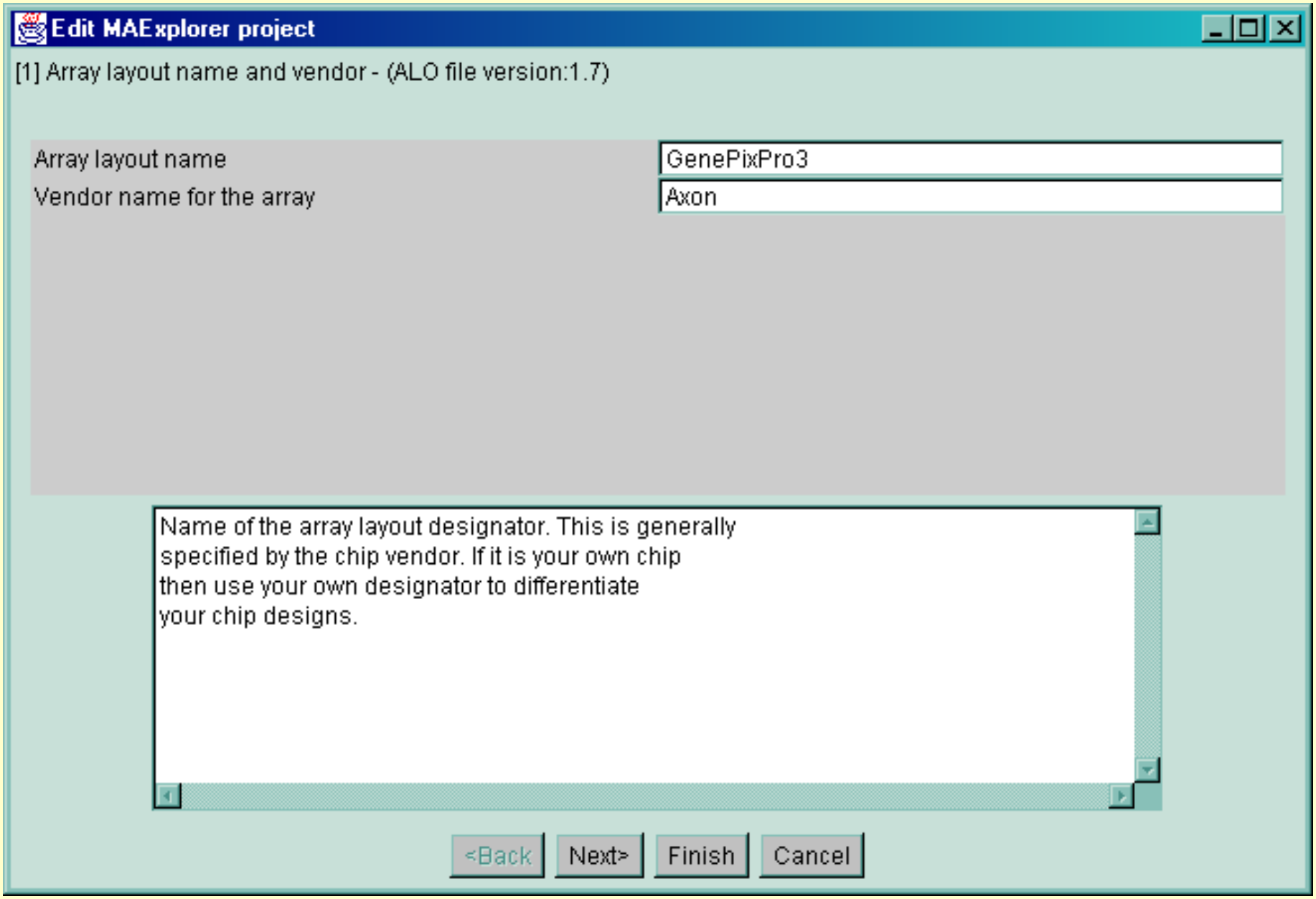
3. Select Files with "Browse input file" Name



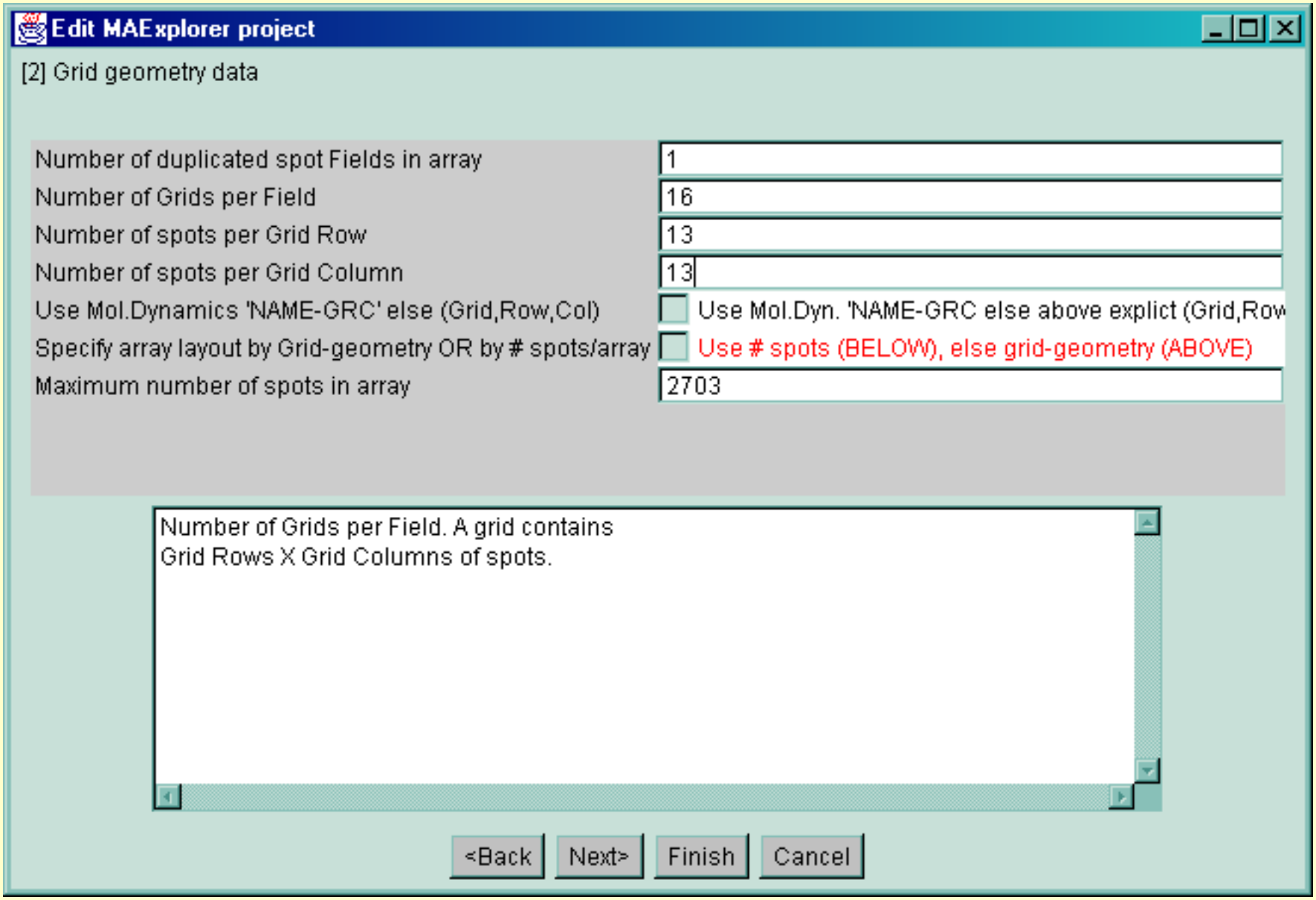
4. Continue Adding Input Files If Needed



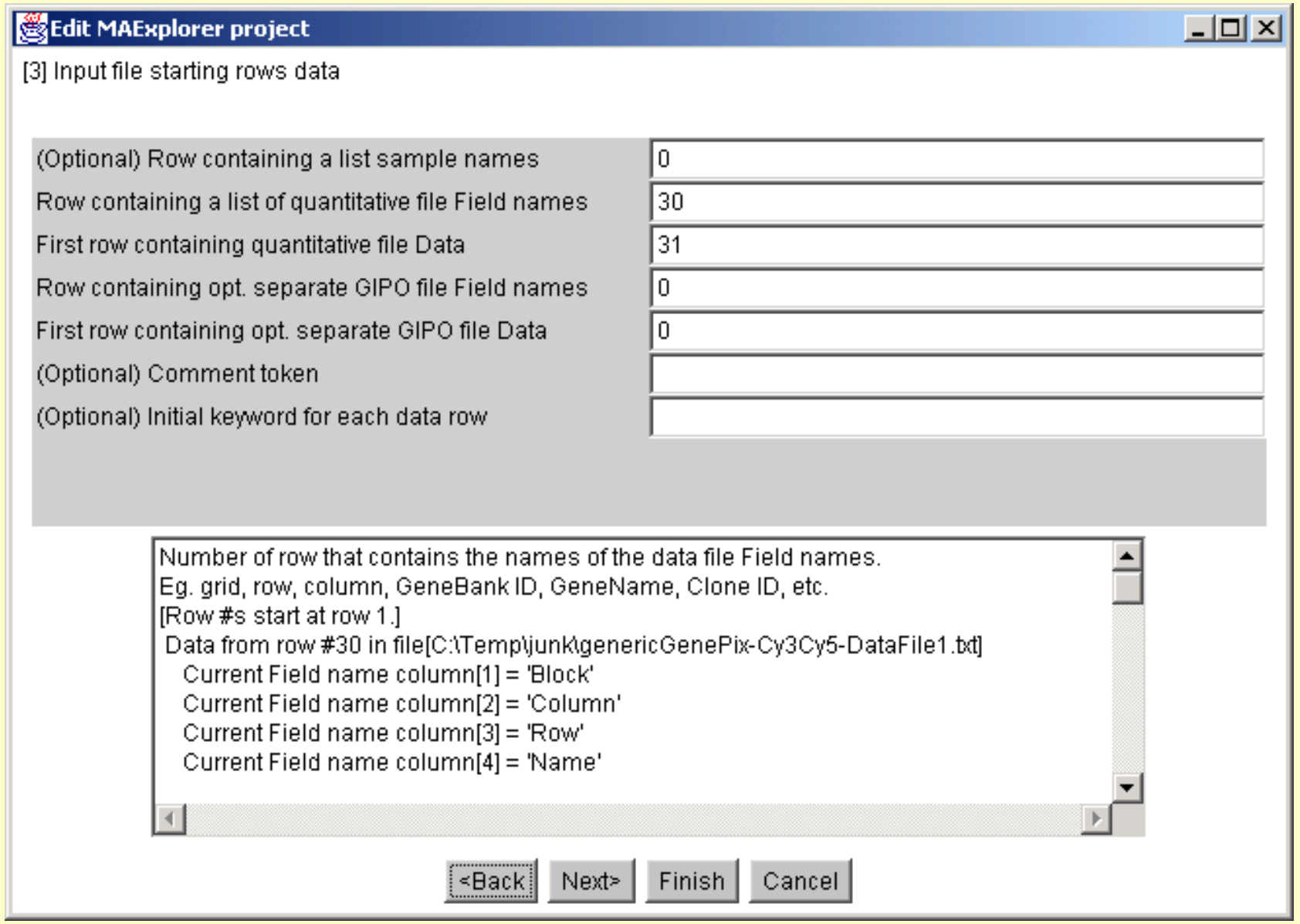
5.1 Edit Layout 'Wizard' Values for This Array



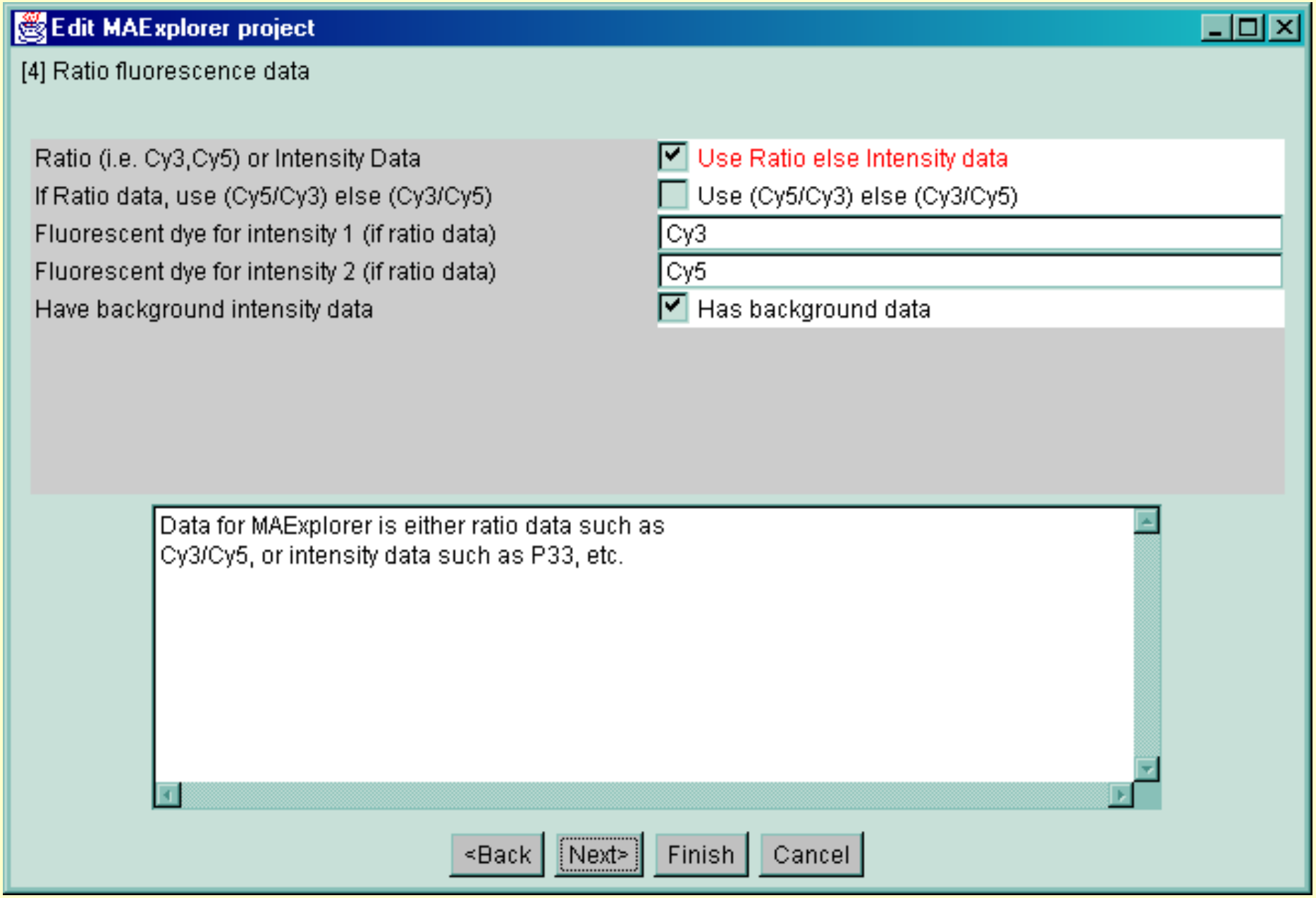
5.2 Edit Layout 'Wizard' - Grid Geometry. Enter (Grid, Rows/Grid, Columns/Grid) Values



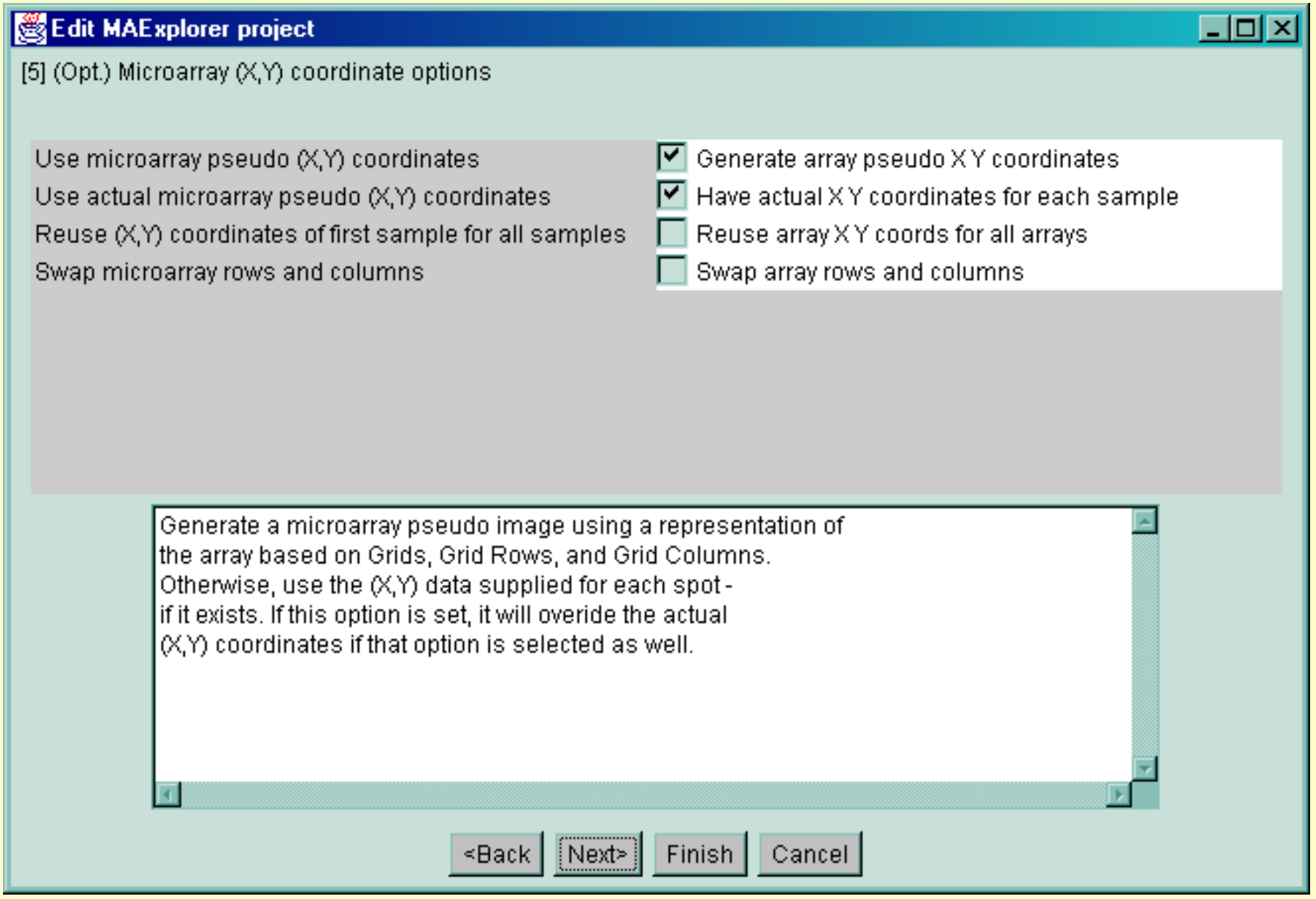
5.3 Edit Layout 'Wizard' Input File Row Values. Verify Row Where Field Names Are Defined



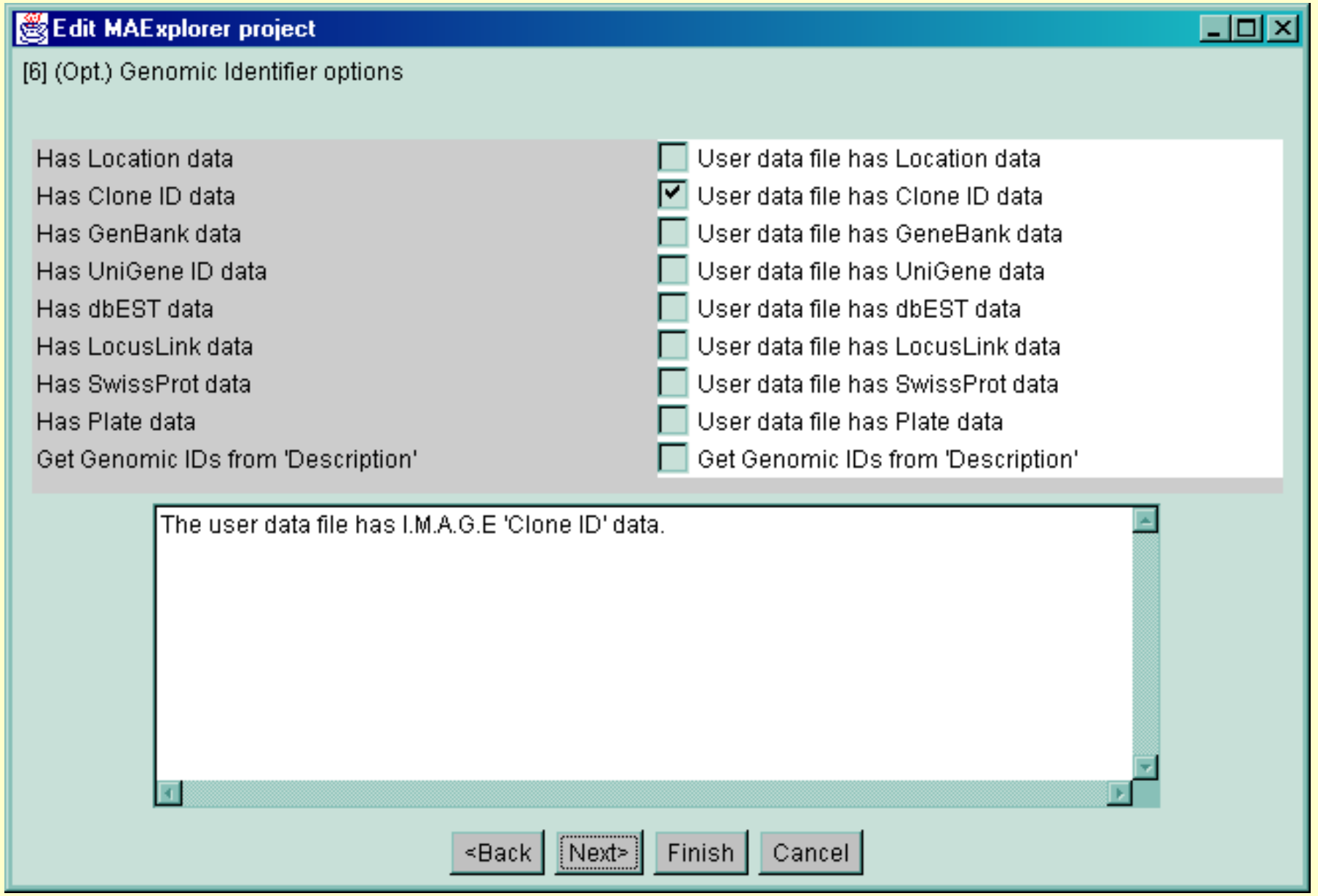
5.4 Edit Layout 'Wizard' Ratio or Intensity Values



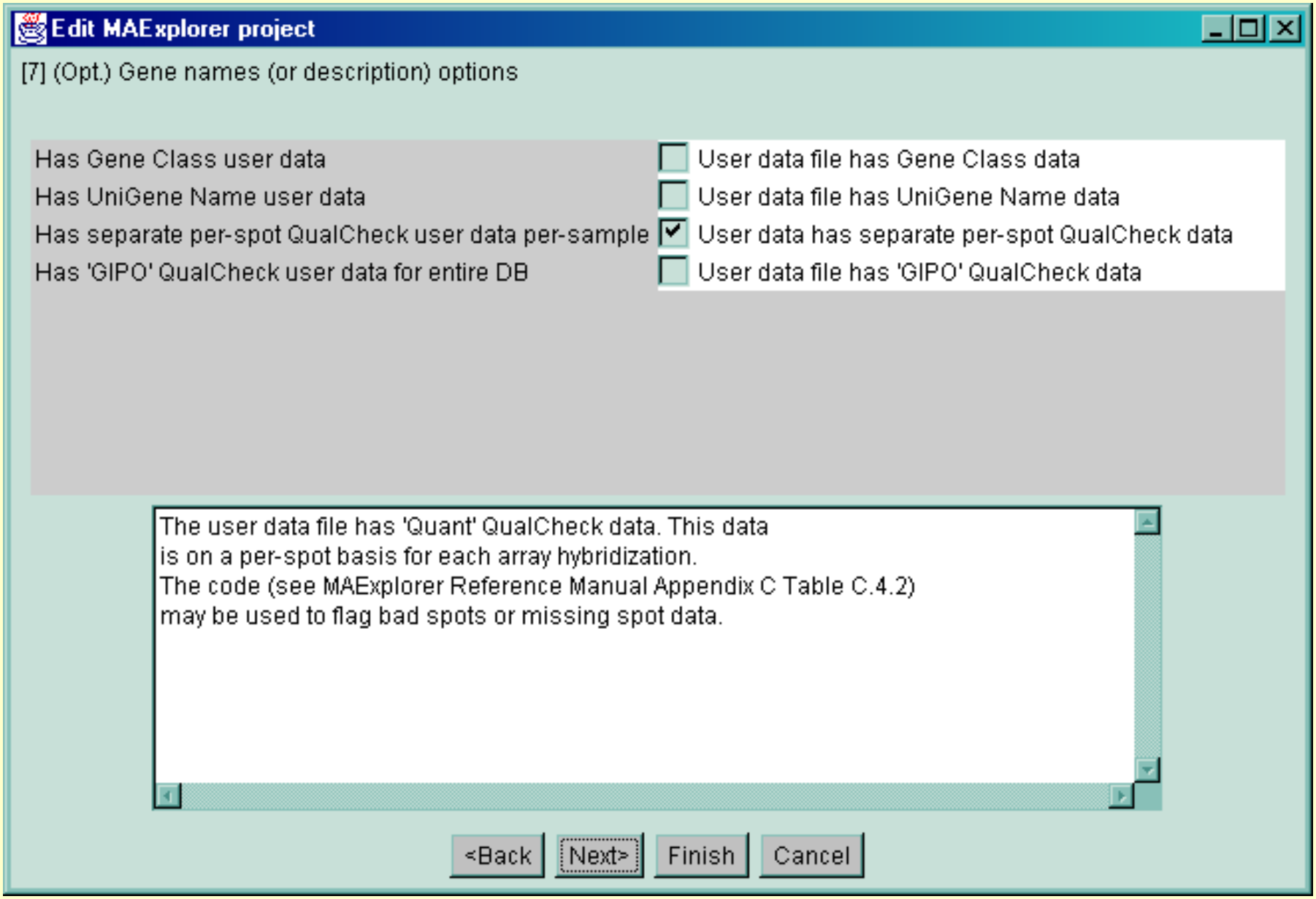
5.5 Edit Layout 'Wizard' optional (X,Y) Coordinate Values



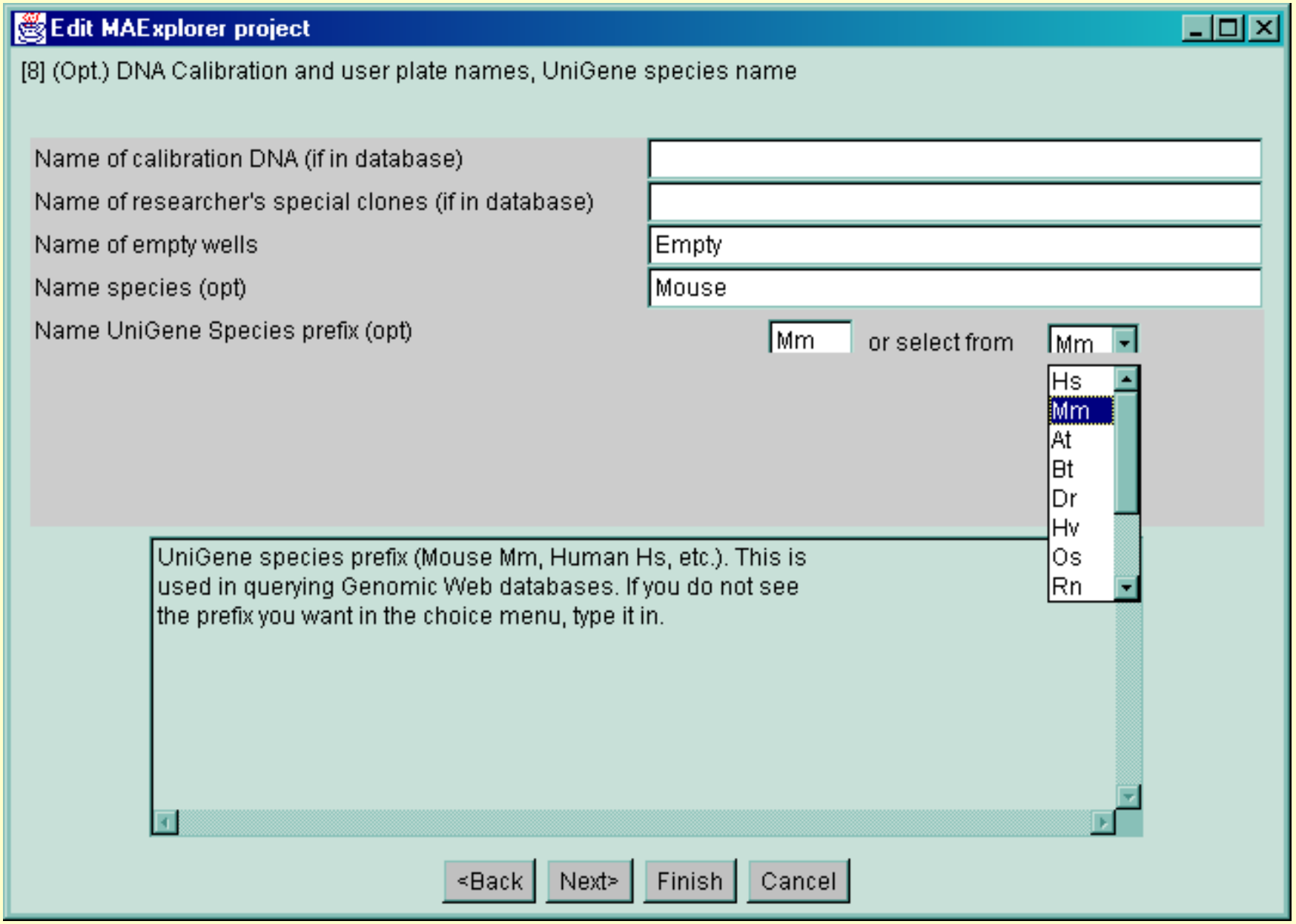
5.6 Edit Layout ‘Wizard’ Genomic ID Values



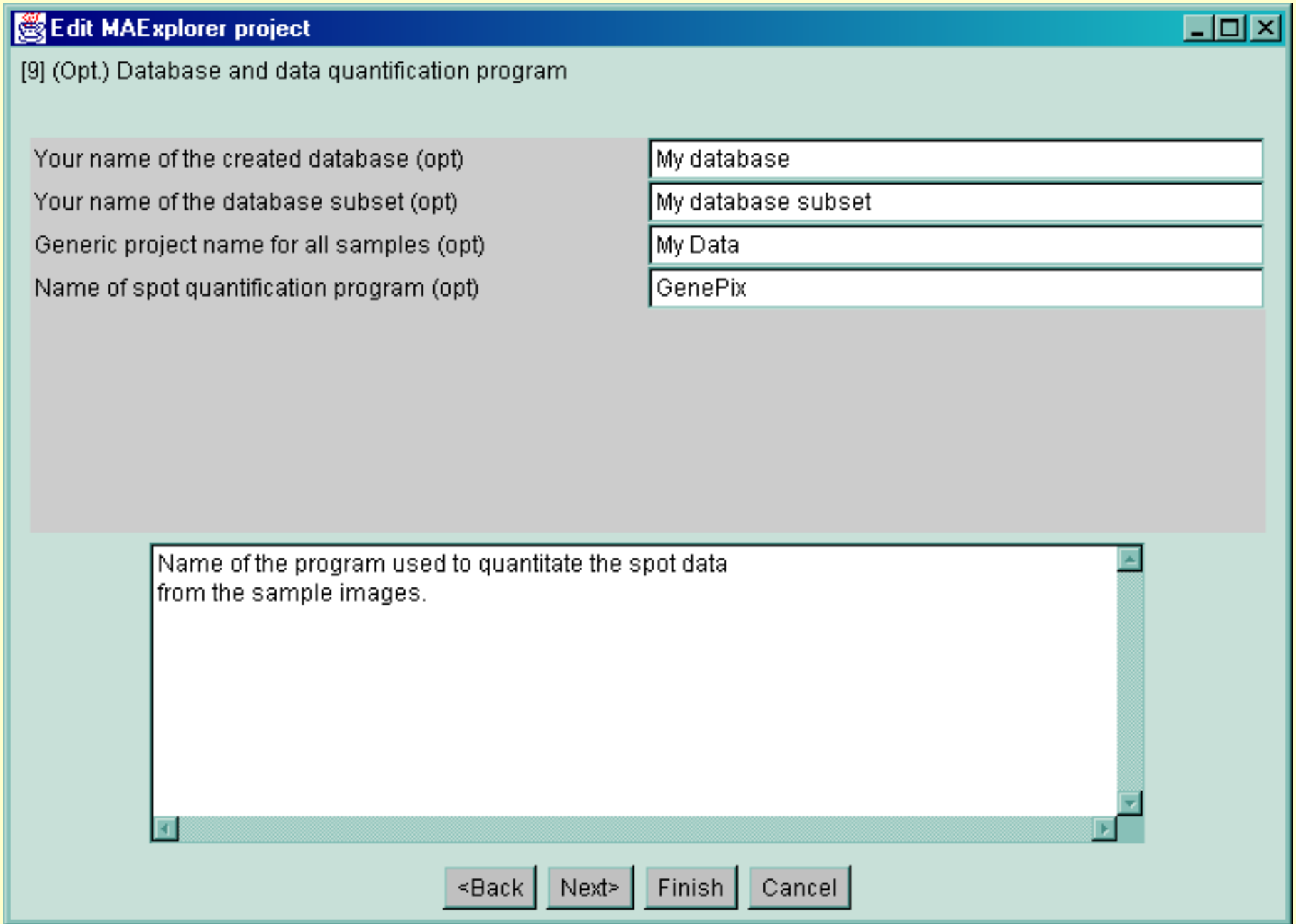
5.7 Edit Layout ‘Wizard’ Gene Names Description



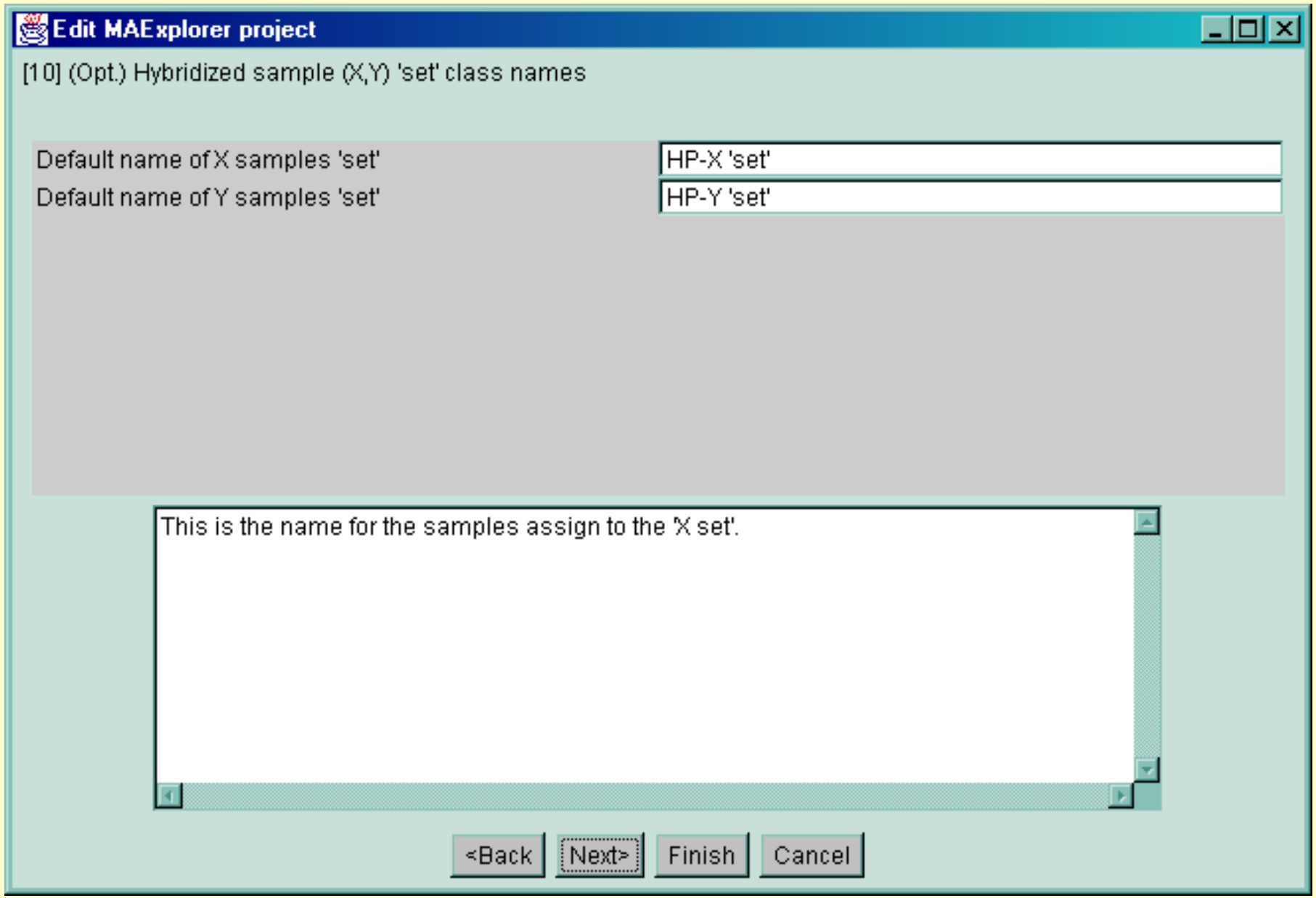
5.8 Edit Layout ‘Wizard’ Calibration Values. Define UniGene Species prefix



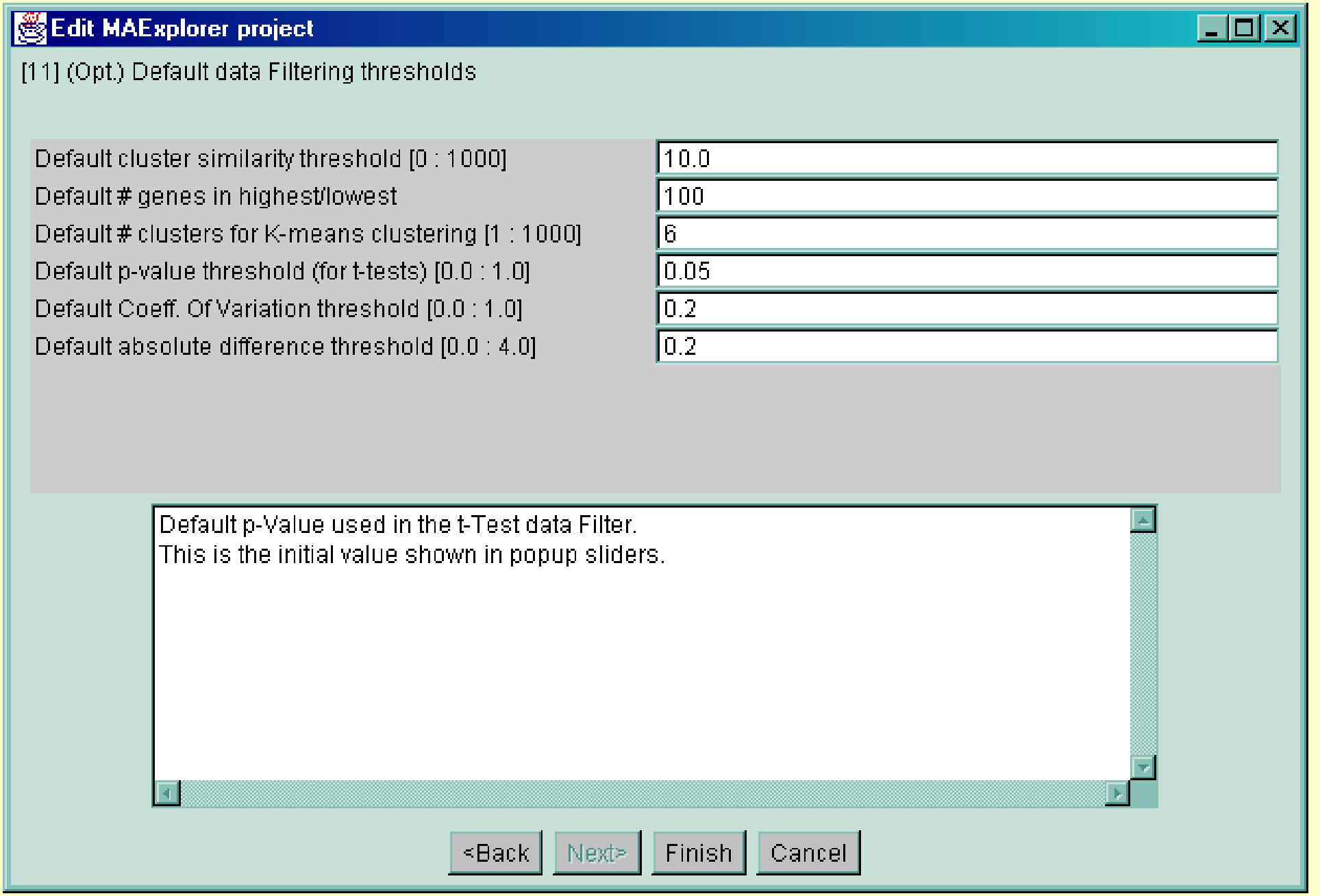
5.9 Edit Layout 'Wizard' Database Name Values. Define Optional Names for Database



5.10 Edit Layout ‘Wizard’ HP-X,-Y Class Names



5.11 Edit Layout 'Wizard' Default Thresholds

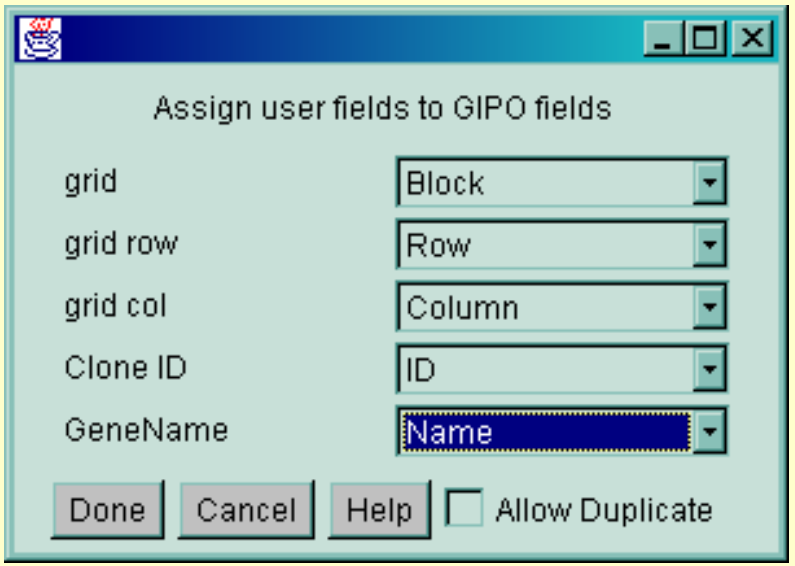


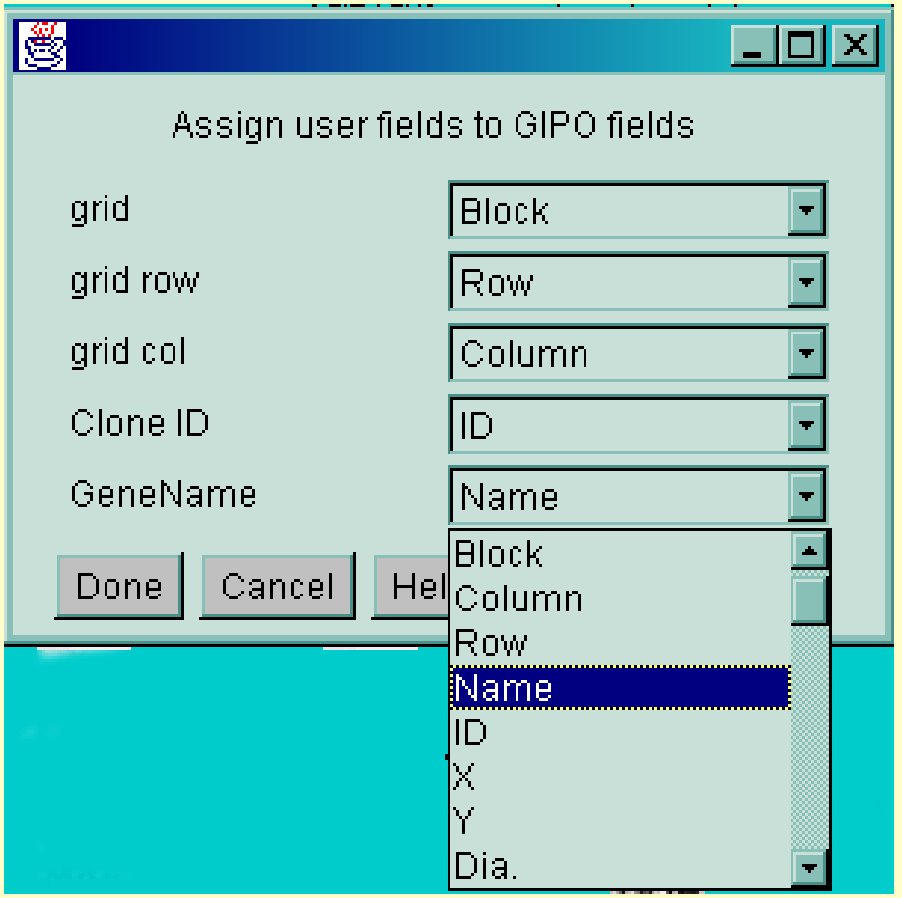
6. Other Options - Assigning User Data Fields to MAExplorer Fields

- GIPO (Gene In Plate Order or "array print" table) - assigns genes to positions on the array as well as GeneBank ID, Clone ID, LocusID (if available), Gene Name, etc.

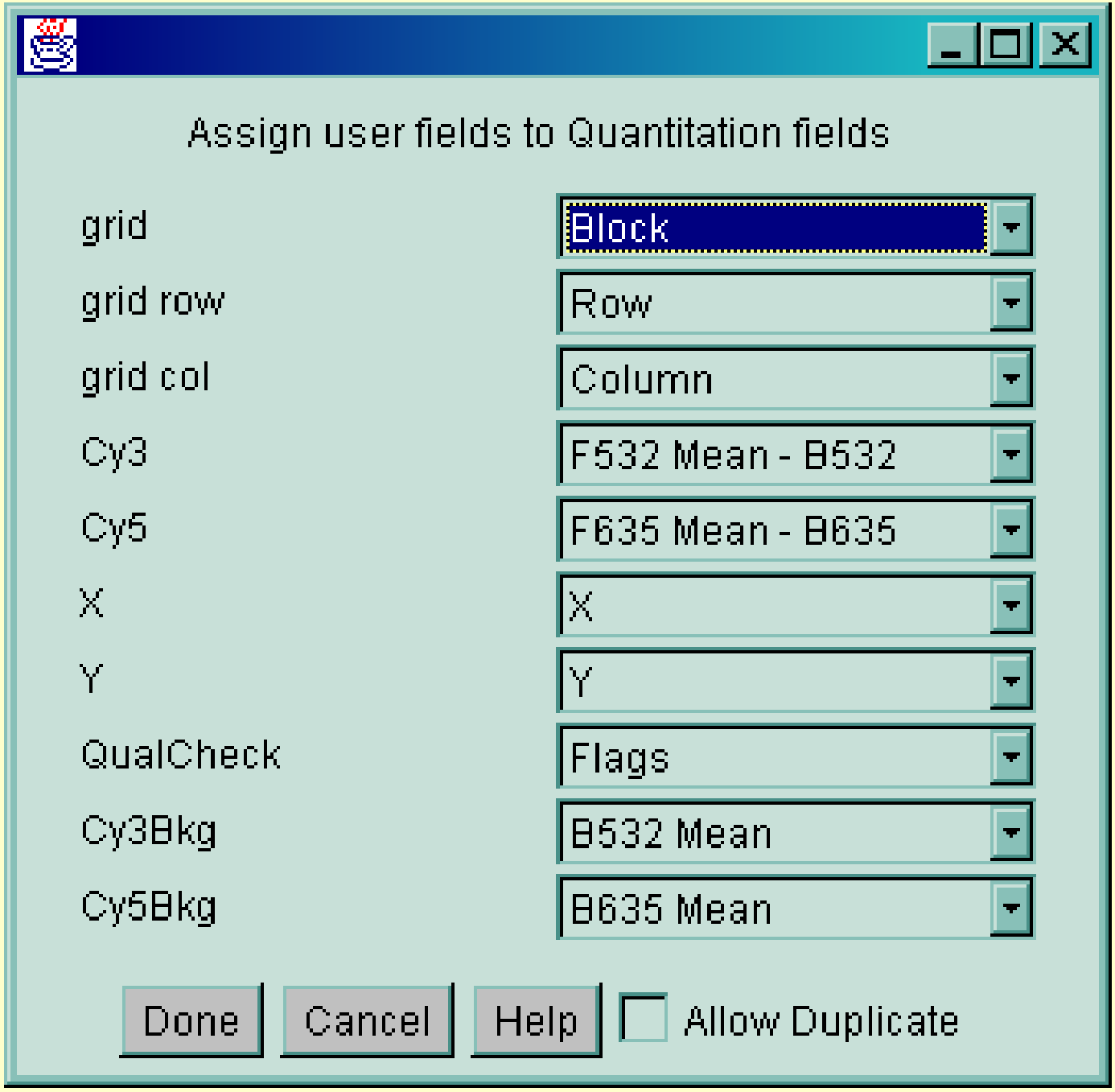
- Quant data - assigns names of quantified data in the user file to MAExplorer data (e.g. Cy3 intensity to RawIntensity1, Cy5 to RawIntensity2, etc).

6.1 “Assign user fields to GIPO fields”

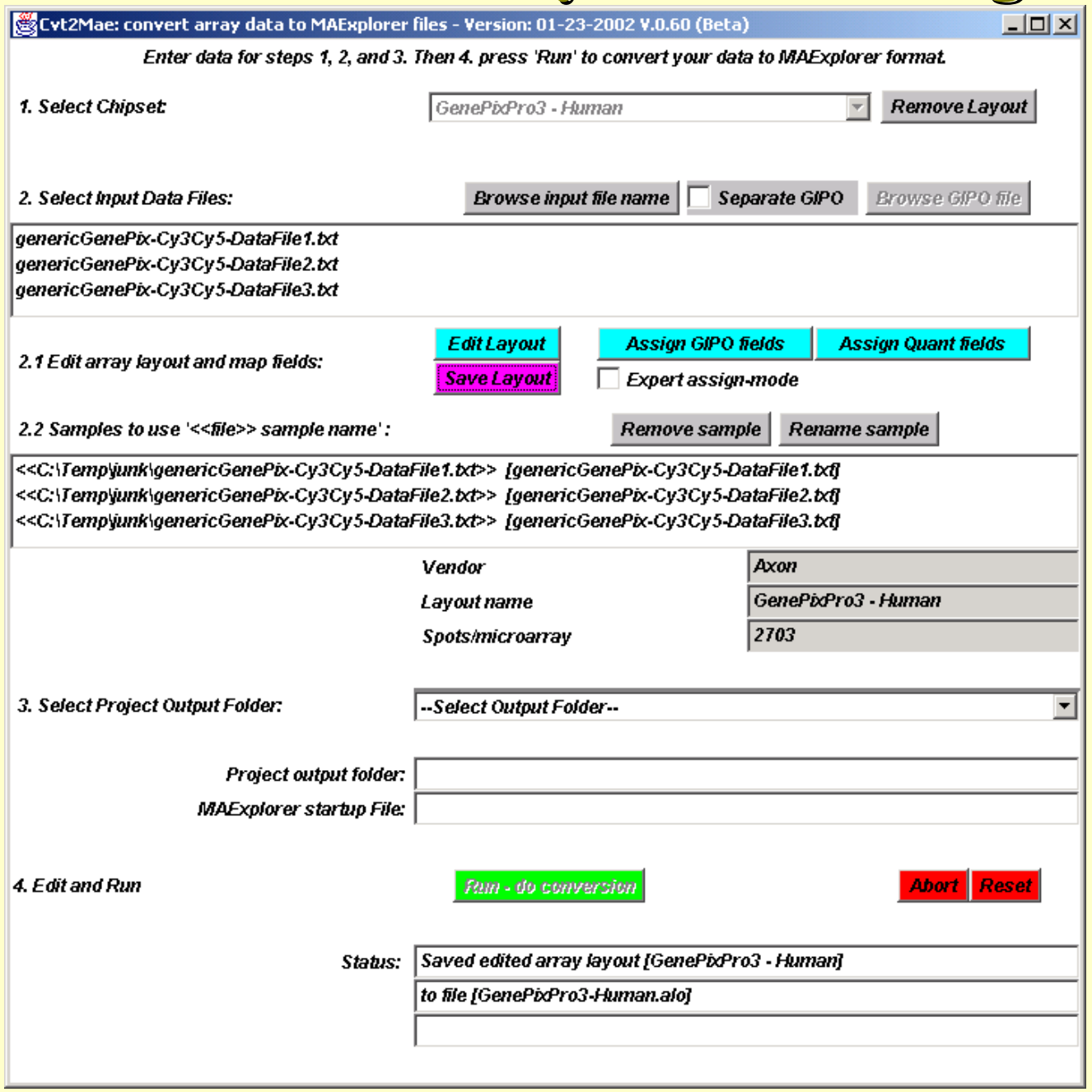




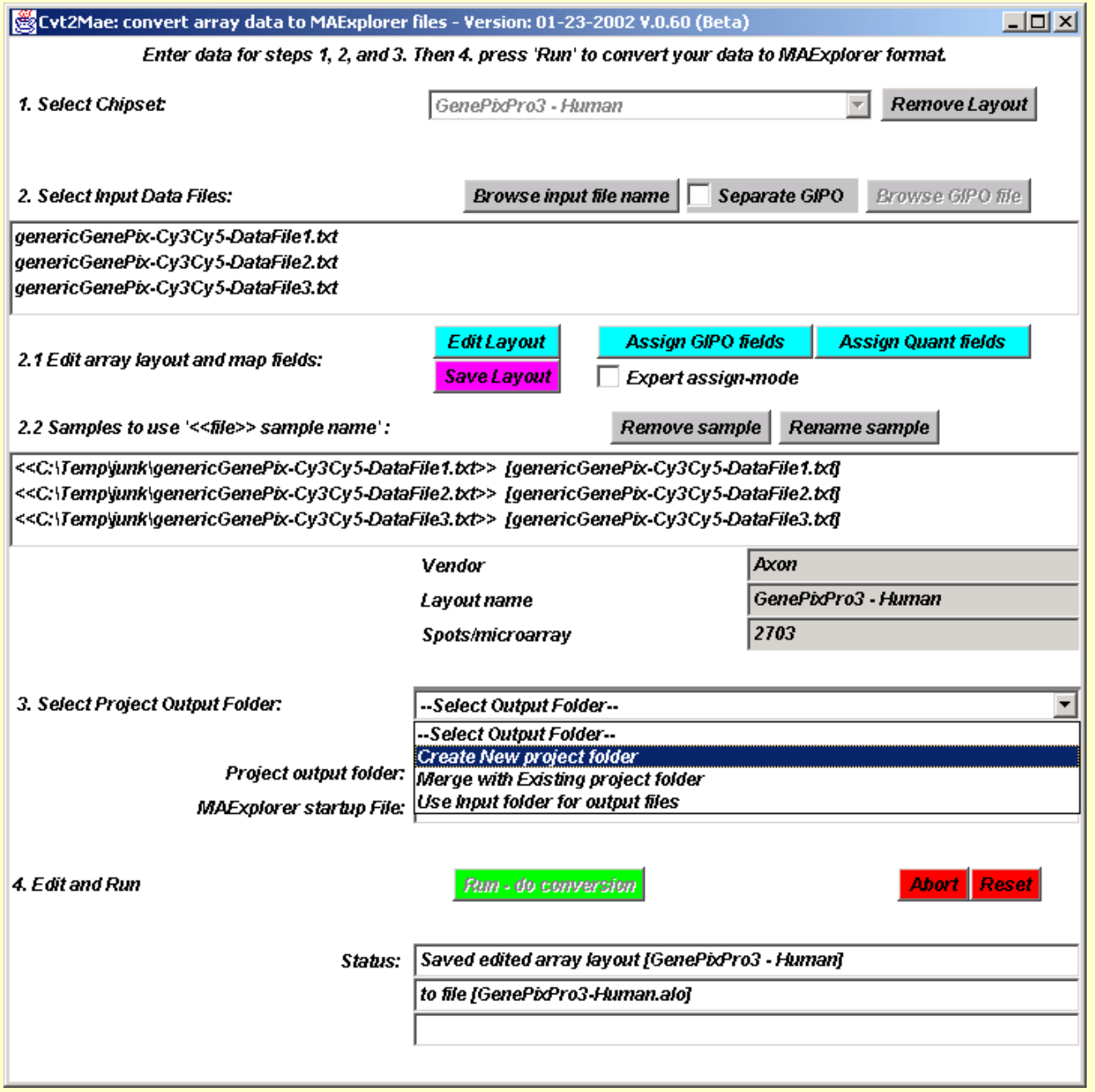
6.2 “Assign user fields to GIPO fields”



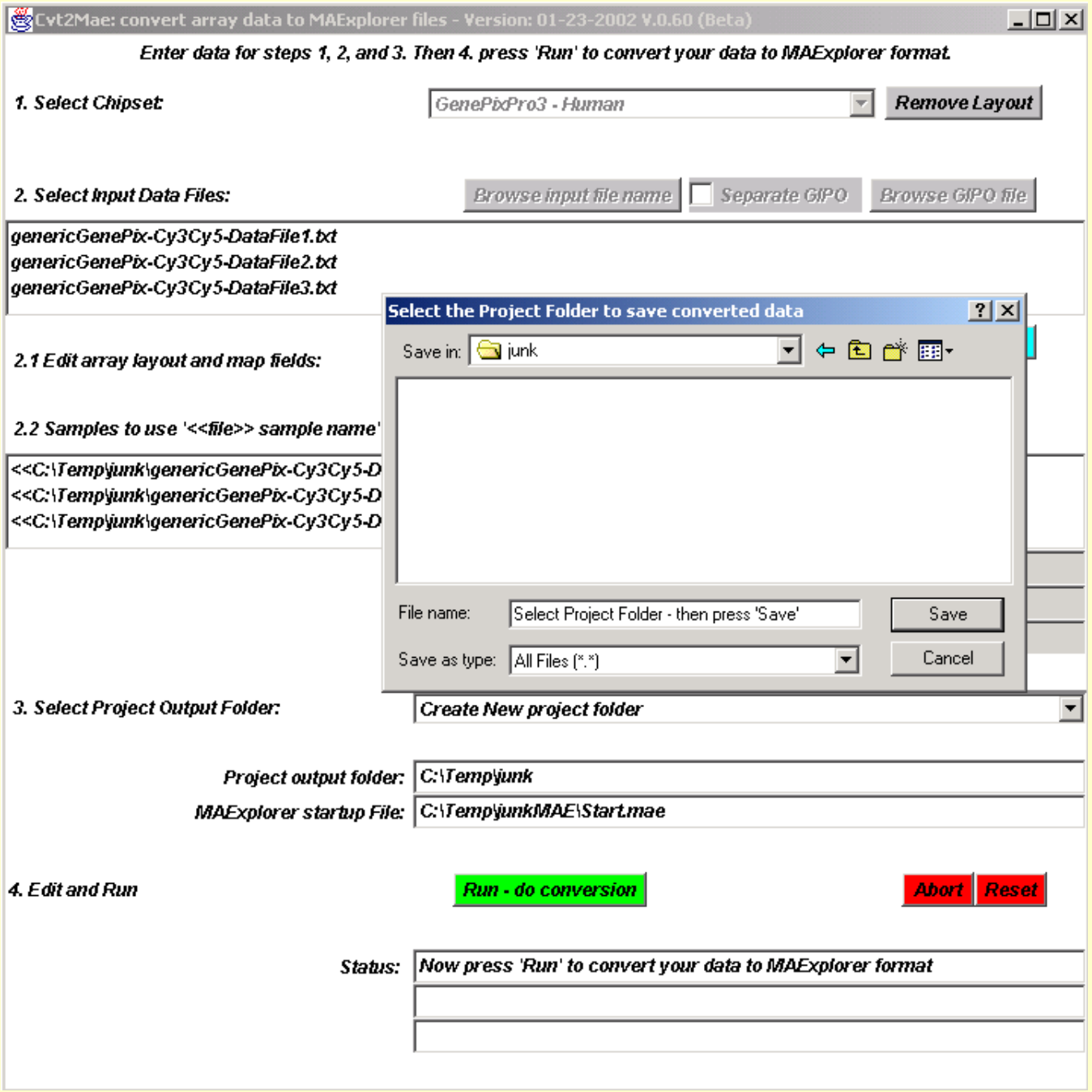
7. Optional "Save Layout" to Array Layout Database After Edit Layout and Assign fields



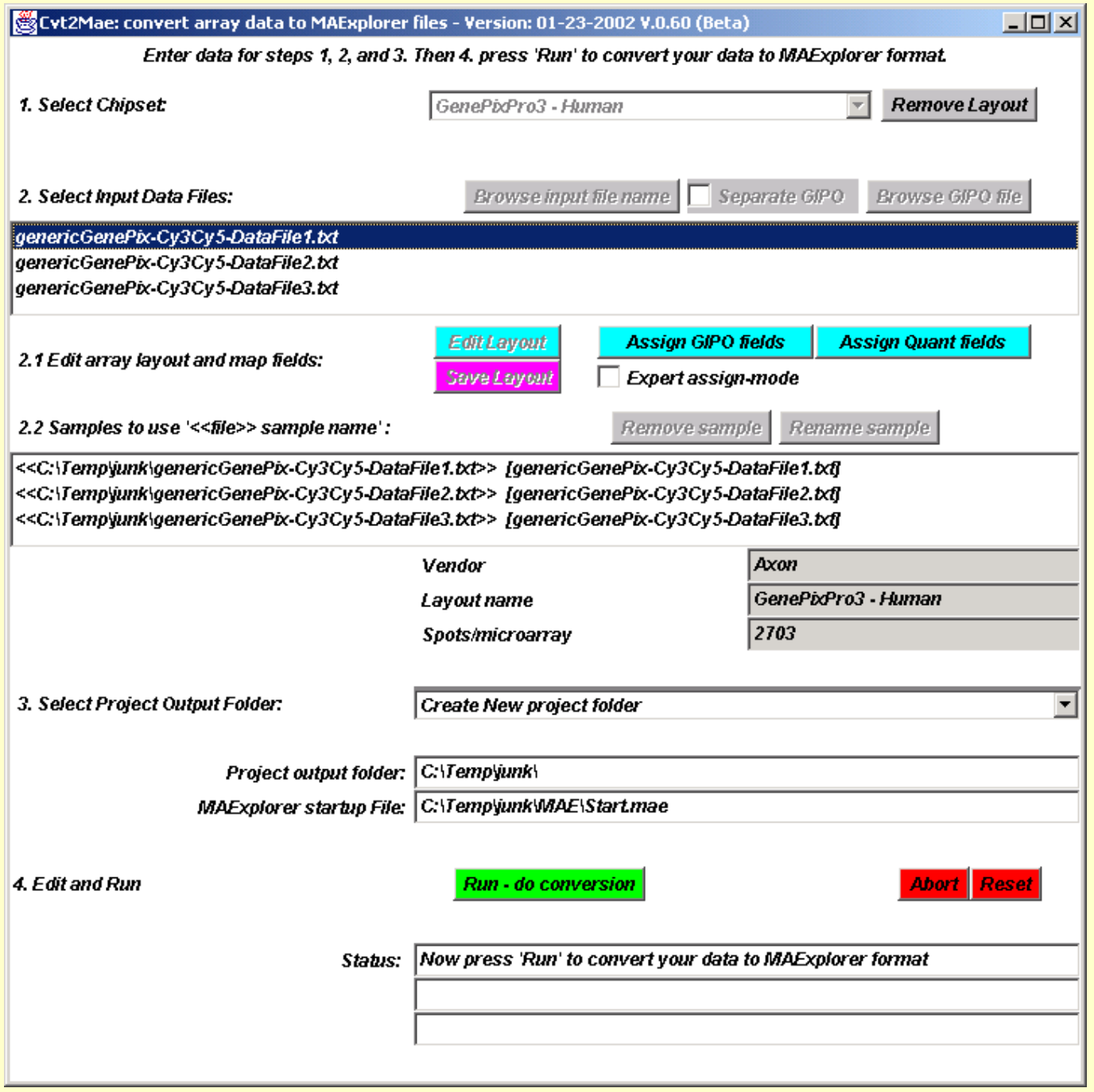
8. Specifying “Create new project folder” Option Where Generated Database Will Be Saved



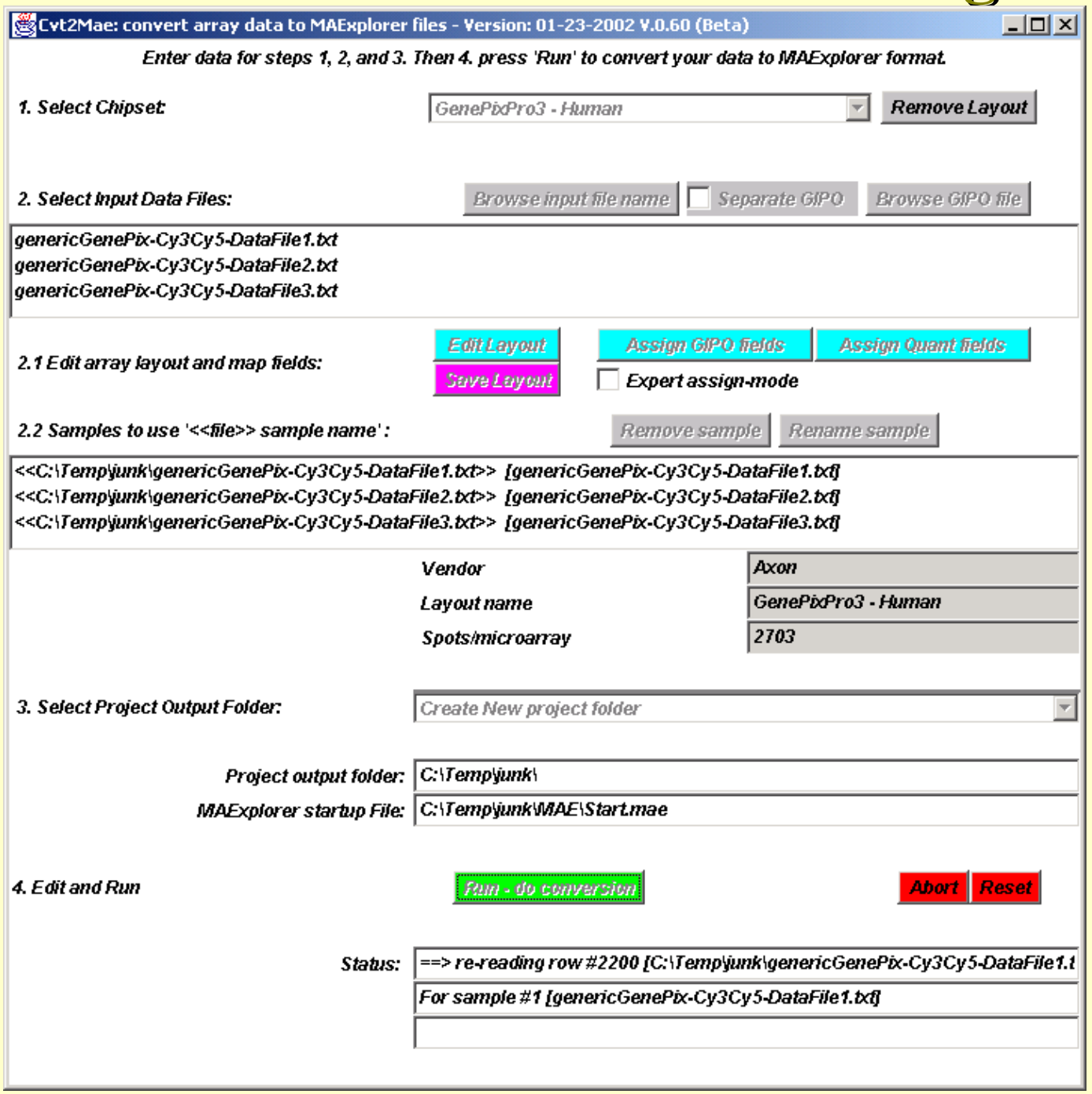
8.1 Specifying New "Project Output Folder"



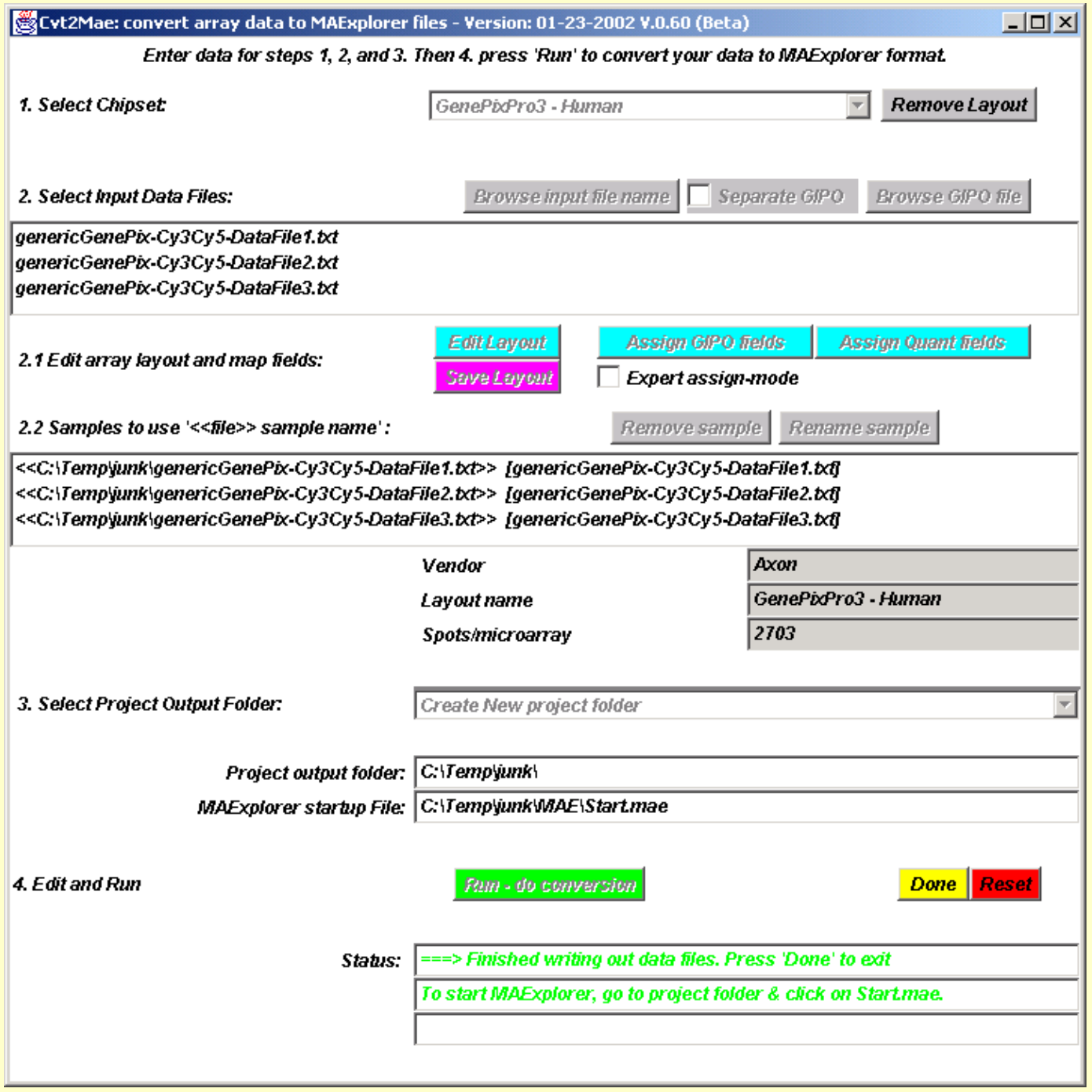
8.2 “Project Output Folder” & MAE startup file



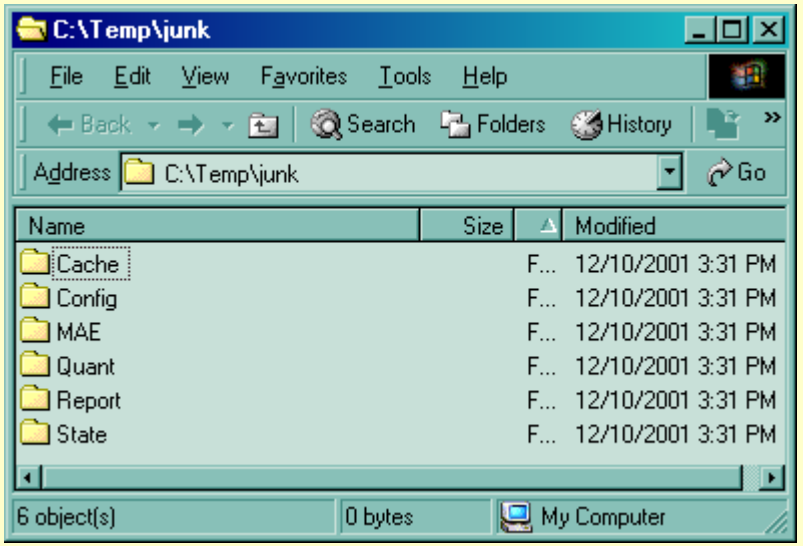
9. Conversion in Process After Pressing "RUN"

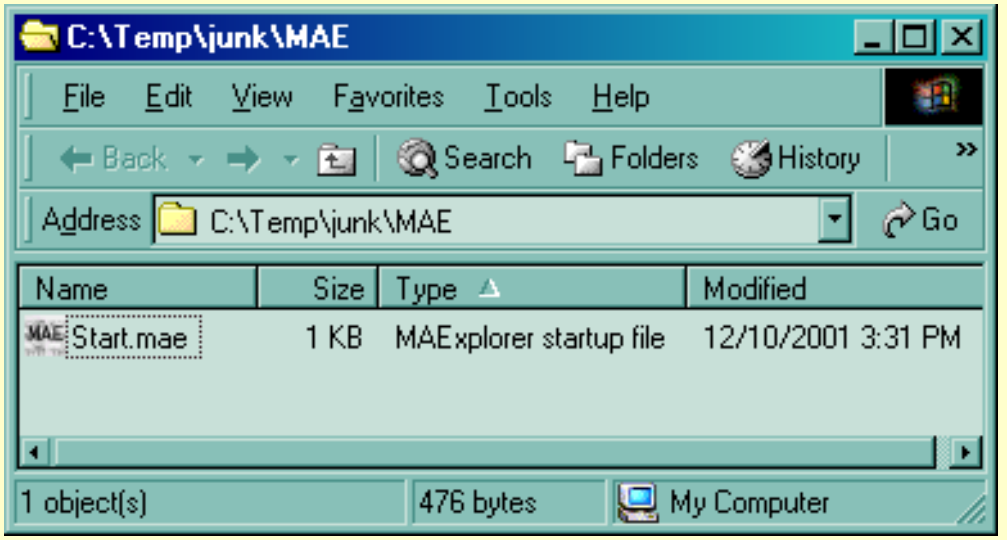


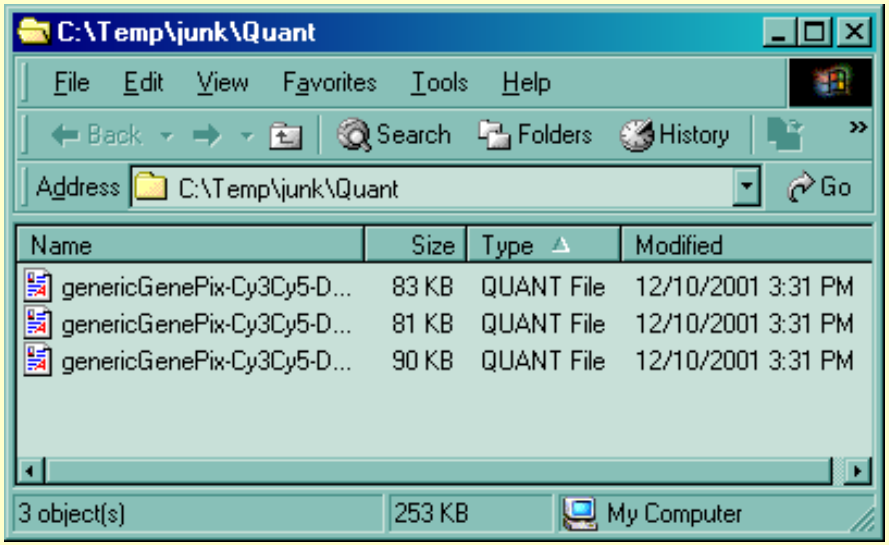
10. Notification that Conversion is Finished

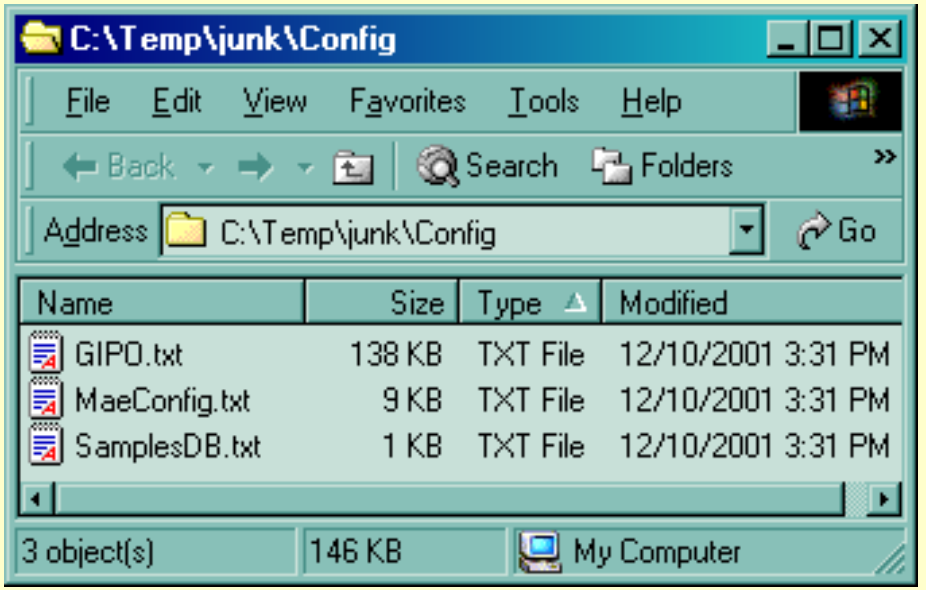


11. MAExplorer Data Created By Cvt2Mae









12. Running MAExplorer on the Converted Data