# Sperm analysis

# Processing measurements files from the Leica microscope

## Preamble

Sperm cell measurements made on the Leica microscope are stored as separate Excel files for each accession/sperm sample (ref. [Bird Collection BPM - Sperm analysis - Sperm cell measurements on the Leica microscope using LAS 4.1](file:///\\lagringshotell\nhm-data\aves\Best%20Practice%20Manuals\Bird%20Collection%20BPM%20-%20Sperm%20analysis%20-%20Sperm%20cell%20measurements%20on%20the%20Leica%20microscope%20using%20LAS%204.1.docx)). This procedure describes how to move or copy these files to a longtime storage location, and to process and prepare the data in the files for inclusion in the [Sperm Morphology Database](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\01%20Sperm%20Morphology%20Database\Sperm%20Morphology%20Database.xlsm) file and futher analyses.

## Preparations

This BPM refers to and makes use of the following folders, all currently located in the folder [Y:\aves\Samlinger\Spermiesamlingen\](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen):

|  |  |
| --- | --- |
| **Folder** | **Procedure(s)** |
| [Morphology DATA files\](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\Morphology%20DATA%20files\) | **A** |
| [Morphology IMAGE files\](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\Morphology%20IMAGE%20files) | **A** |
| [Morphology DATA PROCESSING\Leica measurements files processing\](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\Morphology%20DATA%20PROCESSING) | **B-E** |
| [Morphology RESULTS](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\Morphology%20RESULTS) | **D-E** |
| [01 Sperm Morphology Database\](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\01%20Sperm%20Morphology%20Database) | **F** |

*Procedure E* (*Data quality check*) requires the statistical software package R to be installed on the computer.

## Procedures

### Depositing files for longtime storage

1. Files created using the Leica microscope should be deposited in two different locations, for Excel data files and image and associated files, respectively;
   1. Copy all Excel files from the Leica computer to the correct species folder in Morphology DATA files\
      1. If this does not already exist, create it using the English name as given in Corema as the folder name
      2. The files may well be placed in a subfolder with an informative name (e.g. your name and year, or similar), so that it later will be evident that these files have been measured together
   2. Copy all other files, i.e. excluding the Excel files but including the hidden folder .Metadata and all image files from the relevant folder on the Leica computer to the correct species folder in Morphology IMAGE files\
      1. If this does not already exist, create it using the English name as given in Corema
      2. If the Excel files were placed in a subfolder, do the same with the image and associated files

### Merging individual files

1. Make sure that the folder Files to be merged (in Morphology DATA PROCESSING\Leica measurements files processing\) is empty; if not, delete all files in it
2. Copy all Excel files to be processed together (viz. typically all new files of one species) from the respective folder in Morphology DATA files into the Files to be merged folder
3. Open the file [01 Merge files.xlsm](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\Morphology%20DATA%20PROCESSING\Leica%20measurements%20files%20processing\01%20Merge%20files.xlsm)
4. Make sure that the file type shown in cell B4 matches your measurements files (default is XLS)
5. Click the Merge files button to start merging the files. When all files have been merged – which may take some time! - the resulting file will be shown

NB! It may be smart not to use the computer for anything else while the files are being merged, especially if you are merging a large number of files

1. In the merged file, data from the first sample should appear in rows 1-38, with *Image Name* in cell A1, *Specimen* in A2 etc., and the actual measurement data in rows 9-38 (assuming that 10 cells have been measured). Data from the next sample should then start in row 39, but omitting the first row with *Image Name* so that *Specimen* should appear in cell A39 (this is because the merging macro regards the first row as a header row and therefore includes it only once in the resulting file). Data from the second sample should appear in rows 46-75. However, extra rows are sometimes added between the samples, and these should be removed before proceeding with the procedure! Further, extra rows should be added for samples with fewer cells measured. Therefore, do the following;
   1. Scan through the file and make sure that for every transition to a new sample, the last row of data for one sample is followed directly by a row with the *Specimen* info from the next sample – there should be no empty rows in between here (the total number of rows should be (37 x number of files) +1)! Be aware that there SHALL be two empty rows between the info fields and data within each sample (i.e. after *Calibration* and before the data headers).
   2. If there are samples with fewer cells measured, make sure to add empty rows after the last row of data for that sample so that the total number of “data” rows is identical for all samples in the merged file
2. Close the 01 Merge files.xlsm file and delete the measurements files from the Files to be merged folder
3. Keep the merged file open while proceeding to the next step

### Processing merged file content

1. Open the file [02 Data processing.xlt](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\Morphology%20DATA%20PROCESSING\Leica%20measurements%20files%20processing\02%20Data%20processing.xlt); if a *Security Warning* appears, make sure to enable macros
   1. The data already present in this file are just for illustrative purposes and will be replaced by the real data in the next steps
2. The following steps are also explained in comments to some of the header cells in the file
3. Copy col. A-E from the merged file in step 9 into 02 Data processing.xlt (all remaining steps in *Procedure C* refer to this file)
   1. The formulas in cells G9 and H9 assume that the *Specimen* info is on the format *SpeciesCode\_AccNo\_ItemNo* (e.g. *CyaTen\_12345\_2*), in order to capture the *Accession* and *Item* numbers, respectively. If this is not the case, edit the formulas in cells G9 and H9 so that these are captured correctly (cells G10:H38 refer to G9 and H9 and therefore need no editing).
   2. The formulas are also based on that there are measurements for 10 sperm cells for each sample; if a different number of cells have been measured, make sure to fill down the formulas in cells G38:M38 as far down as needed to cover all rows with data belonging to the first sample in col. A-E, before proceeding with the procedure
4. Duplicate all cells in the range G2:M38 (or further down if more rows were included in step 12b) down these columns so as to cover all rows with data in col. A-E
5. Copy the values from col. G-M (except row 1) into col. O-U, using *Paste Values...-Values*
6. Sort col. O-U by clicking on the Sort 1 button (this will sort these columns on *AccNo*, *ItemNo*, *Image Name*, *Append No* and *Measurement #)*
7. After sorting, copy the formulas in the range W2-AC2 (should already be pre-selected) and paste them into col. W in the first row with data in col. O-U
8. Remove all rows in col. O-AC with only zeros in col. O (using *Delete...* and *Shift cells up*)
9. Copy formulas and blank cells in the “new” range W2:AC4 (i.e. a block of rows 2-4 in these columns) and paste them down to cover all lines in col. O-U
10. Copy the values from col. W-AC (except row 1) into col. AE-AL, using *Paste Values...-Values*
11. Sort col. AE-AL by clicking on the Sort 2 button. This will sort on *AccNo*, *ItemNo* and *Measurement #*)
12. Complete the running number series in col. AL (*SpermNo*) by double-clicking the black dot in the lower right corner of cell AL4
13. Save the file with a file name on the format *Species name – Data processing YYMMDD.xlsx* (e.g. *African blue tit - Data processing 150112.xlsx*) in the same folder as the individual measurement files are located (i.e. in a folder within Morphology DATA files\)
    1. To retain the functionality of the macros in the file, use Save as type = *Excel Macro-Enabled Workbook (\*.xlsm)*
14. Close the merged file from step 9 without saving it

### Summarizing data

1. Open the file [03 Summarizing data.xltm](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\Morphology%20DATA%20PROCESSING\Leica%20measurements%20files%20processing\03%20Summarizing%20data.xltm); if a *Security Warning* appears, make sure to enable macros
2. Copy all data from col. AE-AL (except row 1) in the file saved in step 22 above and paste them (using *Paste Values...-Values*) into col. A-H in the Raw data sheet of 03 Summarizing data.xltm, starting in cell A2
3. Click on the Format table button to format the table and add formulas
4. Copy all data from col. A-H in the Raw data sheet into the same coloumns in the SpermCells sheet
5. Go to the Pivot sheet, right-click within the pivot table and choose Refresh from the context menu
   1. Check that the number of accessions corresponds to the number of files processed; if not, the *Specimen* info is probably wrong in one or more of the source files. This should then be fixed both in the original file(s) in Morphology DATA files and in the data processing file produced in *Procedure C*:
      1. Identify the error(s) and fix the *Specimen* info in the original Excel file(s)
         * NB! The *Image* *Name* fields, both in the header and in the data section of the file(s) should *not* be edited!
      2. Apply the same changes to the relevant *Specimen* fields in col. C in the data processing file saved in step 22 above (or alternatively redo the file merging described in *Procedure B* and replace all content in col. A-E)
      3. Delete all data, except the first two rows containing headers and formulas, in col. O-AL in the data processing file
      4. Return to step 14 above and redo the processing of the merged data
   2. Check that the number of cells per male in col. G (*Sum of N*) is as expected (this should normally be OK, once the previous step has been passed)
6. Copy all values from rows with real data the pivot table (excluding e.g. header rows and rows with *‘(blank)’*) and paste these into cell A2 in the Males sheet (using *Paste Values...-Values*)
7. Click on the Format table button to format the table
8. Fill down the formulas in the range P3:AR3 (including several condensed columns) so as to cover all rows with data in col. A-J. These data are then ready to be added to the [Sperm Morphology Database](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\01%20Sperm%20Morphology%20Database\Sperm%20Morphology%20Database.xlsm) file (see [below](#_Adding_data_to_1))
9. Save the file with a file name on the format *Species name – Morphology data YYMMDD.xlsx* (e.g. *African blue tit - Morphology data 150112.xlsx*) in the relevant species folder within Morphology RESULTS\
   1. To retain the functionality of the macros in the file, use Save as type = *Excel Macro-Enabled Workbook (\*.xlsm)*
10. Delete col. K-AR
11. Save the Males sheet as a \*.TXT file for use in R analyses in the same folder:
    1. Choose Save As… and select Save as type = *Text (Tab delimited) (\*.txt)*
    2. Enter a file name on the format *SpeciesCode\_Males\_YYMMDD* (e.g. *CyaTen\_Males\_150401*)
    3. Click Save and confirm the two next questions to save the file
12. Save the SpermCells sheet in the same way, replacing *Males* with *Cells* in the file name (e.g. *CyaTen\_Cells\_150401*)
13. Close the file without saving (it was already saved above)

### Data quality check

1. Copy the file [04 Data checking.r](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\Morphology%20DATA%20PROCESSING\Leica%20measurements%20files%20processing\04%20Data%20checking.r) to the folder where the morphology data file was saved in step 32 above, and rename it according to the pattern *Species code\_YYMMDD.r* (e.g. *CyaTen\_150112.r*)
2. Edit the strings within the quotation marks in rows 9 (wDir) and 10 (datafile) to correspond to the folder and file names from step 32 and 35 above, respectively
3. Replace all occurrences of *PruCol* with the relevant species code (using the menu option *Search – Replace* or *Ctrl+R*)
4. Run all code in lines 9-23 and check that the summary info shown in the R window after this corresponds to your expectations
5. Run the code in lines 31-38 to produce four plots, one for each of head, midpiece, tail and total length, where the individual values for each separate accession number are plotted on separate horizontal lines
6. If the plot reveals deviating data points, the code in lines 42-56 can be run to allow labelling of outlying data points in the sub-plots;
   1. Start with running lines 42-44
   2. For the sub-plots you want to annotate, also run the code line starting with identify (e.g. line 45) after the plot line
   3. After the identify code line has been run, labels with the *SpermNo* from the data file can be added to any data point in that sub-plot by clicking close to the relevant point(s)
   4. After all interesting points have been labelled in that sub-plot, right-click in the plot frame and choose *Stop* before running the plot line for the next sub-plot
   5. Repeat steps b-d for all sub.plots you want to annotate
7. Inspect the plots to identify spurious data; typical issues include:
   1. Components have swapped place in the data file; corresponding deviating points will show up in the two plots (e.g. a very long head matching a very short midpiece)
   2. Extreme single values compared to the range for the species or individual
8. If errors are detected, these should be fixed in the original files and the procedure redone from the relevant step
9. The plots may be saved using e.g. the *File-Save as-Jpeg* menu in the plot window
10. Save the .r file before closing it

### Adding data to Sperm Morphology Database (SMD)

1. Open the file [Sperm Morphology Database](file:///\\lagringshotell\nhm-data\aves\Samlinger\Spermiesamlingen\01%20Sperm%20Morphology%20Database\Sperm%20Morphology%20Database.xlsm); be sure NOT to open it in read-only mode, i.e. click the *No* button in the dialog box shown during opening of the file
   1. If the *File in use* dialog box keep popping up, saying that the file is locked for editing, try to open the file from within Excel (*File-Open* etc.); if the same message still appears, the file is really locked for editing, i.e. in use by someone else
   2. If a *Security Warning* is shown at the top saying that *Macros have been diabled*, make sure to click the Enable Content button
2. If you have not used this file before, please take the time to read through the Info sheet, where also the following procedure is described
3. First, unless you are *certain* that the data you are about to add are not already in the SMD (e.g. for samples just added to the collection), go to the Check data sheet and use one of the tables there to check some ID from the new data against all data registered in the SMD;
   1. To check e.g. accession numbers, simply copy the accession numbers from the *Males* sheet in the file saved in step 32 above (col. P) and paste them (using *Paste Values...-Values*) into cell A9 in the Check data sheet
   2. If any of the accession numbers are already registered in SMD, the registered head lengths and the measurer will be shown in col. D-E (or corresponding files if one of the other IDs is used)
   3. By pasting the head lengths from your file (col. AM) into cell B9, also the difference between the new and the already registered head lengths will be shown in col. C and these cells will turn green if the deviance is less than 0.01 µm or red otherwise
   4. The results of the lookups should be interpreted as follows;
      1. If only “*#N/A*”s are shown in col. D-E no data exist in the SMD for any of the accessions
         * Proceed with the import
      2. If data do show up in these columns, but one or more of the cells in col. C turn red, data for the relevant accessions do exist in the SMD but are likely to be from a different measuring session
         * Proceed with the import
      3. If data do show up in these columns and all cells with data in col. C turn green, the exact same data do most likely already exist in the SMD
         * Compare your new data with those already in the SMD more closely to determine whether they are indeed duplicates; if so, all duplicates should be removed from your import data set before proceeding with the import of the remaining data (if any are left)
4. After having passed the previous step, continue to the Main sheet, click the SORT LIST button and then click *Unprotect sheet* on the *Review* tab to enable editing of the sheet
5. Go to the bottom of the table
6. Copy all data from col. P-AV (except row 1) in the summary file saved in step 32 above and paste these into col. B (*AccNo*) in the first empty row below the table in Main using *Paste Values...-Values*
7. Add running numbers in col. A (*#*), starting with the value shown in the originally last, green row (this will always show the next number to be used, after the list has been sorted). Temporarily the last of the newly added rows will then be coloured green, but this will change as soon as the file is sorted again
8. Add the current date to all new rows in col. D (*Date added*)
9. Fill in the *Priority* column (col. E);
   1. If no data exist in the SMD for an accession (ref. step 49 above), set *Priority* = 1
   2. If data do exist for an accession, confer these and decide which one registration (row) shall be regarded as the prevailing measurement (*Priority* = 1); the rest shall have *Priority* = 2
10. Fill in the full name of the person who has made the measurements in col. AB (*Measurer*)
11. Fill in the full file name (incl. file extension) of the summary file saved in step 32 above in col. AJ (*Source summary file*) for all the new rows of data;
    1. If this was saved directly in the species-specific folder in *Morphology RESULTS*, nothing more needs to be done
    2. If it was saved either in a subfolder of the species-specific folder, or in some other folder within *Morphology RESULTS*, paste the full path of that befolder below the *Morphology RESULTS* level into col. AH (*Source summary folder manual*), e.g. *African blue tit\2014* or *Cyanistes*
12. Similarily, if individual measurement files have *not* been named according to the standard schema (viz. *CyaTen\_12345\_2*.xls) and/or *not* placed in the species-specific folder in *Morphology DATA files*, this will have to be filled in manually for all the new rows of data;
    1. If placed in a different folder, fill in col. AL (*Ind. file subfolder manual*) as explained in step 57b above
    2. If named otherwise, fill in the full file name (incl. file extension) of the file in col. AM (*Ind. file manual*)
13. Fill down the formulas from all the the gray cells in the originally last, green row (col. F-Y, AK and AN-AY) to all the new rows of data. Check the looked up data in these columns to see that they seem reasonable and correct;
    1. If not – find out why and fix it!
    2. If no data (#N/A) are shown, the background data needs to be updated:
       1. Export an Accessions - Data for Sperm Morphology Database report from Corema for all accessions with *Sperm*, *Testes* and/or *Seminal glomera* items (by selecting the saved filter Accession data for Sperm Morphology Database these criteria will be automatically applied)
       2. Replace all data in the Corema Acc. data sheet with the newly exported Corema data
       3. Export an Items – Sample data for Sperm Morphology Database report for all *Sperm*, *Testes* and/or *Seminal glomera* items (by selecting the saved filter Sample item data for Sperm Morphology Database these criteria will be automatically applied)
       4. Replace the data in col. B-F in the Corema Sample item data sheet and fill down the formula in col. A as far down as needed to cover all rows with data in col. B-F
       5. Export an Items – Slide data for Sperm Morphology Database report for all *Existing Sperm slide* items (by selecting the saved filter Sperm slide item data for Sperm Morphology Database these criteria will be automatically applied)
       6. Replace all data in the Corema Slide item data sheet with the newly exported Corema data
14. Convert the formulas in the columns with green headers (I-Y and AP-AY) to text by copying their content and pasting it directly back using *Paste Special... – Values*
    1. NB! Leave the formulas in the coloumns with *blue* headers (F-H, AK and AN-AO)!
15. Sort the list (button SORT LIST) and check the new data (e.g. by searching for the current date in col. D) to make sure everything seems OK - if not; fix it! :-)
16. After all data have been added and checked, go to the Review tab, click the Protect Sheet button and click OK **without entering any password**.
17. Go to the Info sheet and update the *Last updated* date
18. Save and close the file