1. **Interpretation of metagenomics results *(By Patrick Munk, Technical University of Denmark; pmun@food.dtu.dk)***

*The objectives of this section are to explore read count matrices based on metagenomic results, calculate how similar samples are, visualize abundance matrices using PCoA plots and heatmaps, and determine whether an explanatory variable influences the resistome composition.*

In this section, we will provide you with tables containing the matrices derived from mapping reads from the pig and broiler samples metagenomes to the ResFinder database.

**The first step is to download the .zip file section IV material. Afterwards, make sure you extract the individual files into the Vboxshare folder. The files included are the four tables described below and the files Exercise IV read count analysis.Rmd and WorkshopRfunctions.R :**

EFFORT\_MG\_sampledata\_final.txt - A table that describes the farm each of the samples originate from

EFFORT\_MG\_ResFinder\_count.txt - The count table with reads matching ResFinder genes in the samples

EFFORT\_MG\_ResFinder\_FPKM.txt - A gene-level relative abundance table for AMR (FPKM)

EFFORT\_MG\_ResFinder\_class\_FPKM.txt - A class-level relative abundance table for AMR (FPKM)

Next, start Rstudio **from the terminal in the virtual machine with administrative rights ("sudo rstudio").** Now open the script (R markdown file) Exercise IV read count analysis.Rmd. Note that it might not work if you launch Rstudio from the desktop icon!

Follow the script steps, running code “chunks” one by one (*press the forward green arrow on top of each code chunck – if you stand over the arrow, you should read “run current chunck”*). Read the explanations in between the code to understand the analysis.

For convenience, you can reply to the questions on the script – these are the quiz questions for section IV!

**NB. You can save your replies to the questions on the script itself, by writing your reply outside of areas of code chunks.**

QUESTIONS:

1 - What is the range of different AMR genes observed in the samples?

2 - Do you think these values are meaningful to compare?

3 - Does host animal appear important to the structure of the resistome?

4 - How much of the resistome variance is explainable when using only two dimensions?

5 - When you modify your last command to color based on country, do you see a country effect?

6 - Does it look like there is a country effect on the pig resistome?

7 - Is the country effect significant at alpha 0.05?

8 - By changing the last code line, now do the same analysis for the pig data (variable: "RF.gene.fpkm.pig"; table "sampledata.pig") and color-annotate by "country". In which country is the country effect (clustering) less pronounced?