

Homework Week 2

Friederike Dündar and Luce Skrabanek

ANGSD Course 2020

Please address the following questions and exercises using an `.Rmd` document that states your name and the date at the top.

Relevance of DNA sequencing (7pts total)

1. Why are we actually interested in the order of the DNA's base pairs? Give an example application of DNA sequencing and its aim/relevance. **3pts**
2. Explain two differences between the traditional Sanger and next-generation-sequencing. **2pts**
3. Find a publication that uses RNA-seq data (tell us how you found it, too). Identify the main question that is being addressed with the RNA-seq data in that paper. **(2pts)**

Exercises (7 pts total)

1. Make a folder in which you're going to keep track of everything related to your homework for the ANGSD class. (*UNIX*) **1pt**
2. Download the files containing the **lengths** of the individual yeast (*S. cerevisiae*) **chromosomes** for every genome assembly that is available at the [UCSC Browser](#) (hint: these should be very small text files ending with "chrom.sizes"; you can find them via the links to "Genome Sequence Files" once you've tracked down the site for the yeast genomes). (*UNIX*) **2pts**
3. Compare the files of the different assemblies. What do you notice? (*UNIX*, common sense) **3pts**
4. Make a table listing the sizes for every chromosome across the 3 different assemblies. (*Rmarkdown*) **1pt**

Make sure to document *every* step of your analysis that is relevant to the final results you're presenting (including trivial steps such as checking the content of a file). It's fine to also note if you manually copied and pasted entries, but we (and your future self) want to know exactly how you obtained every piece of information and how any visualizations or summaries were achieved.

Compile the .Rmd file and send both the .Rmd and the HTML files to angsd_wmc@zohomail.com by Saturday night.

Get git talking to RStudio.

Make use of Mervin's office hours!!

1. Work through the git tutorial at <https://happygitwithr.com/>. At a bare minimum, get Rstudio talking to git on your local machine; try to get RStudio talking to GitHub, i.e. chapters 6-13.