

----- BEGIN DO NOT EDIT -----

Please write:

- Study (with link) you are planning to use
- biological question you are planning to answer
- Basic steps
- Names + IDs

Example:

Group 0

Study: <https://www.sciencedirect.com/science/article/abs/pii/S0950268819300940>

Question: Was the bacteriophage ϕ X174 created by aliens?

steps:

1. Quality control of the raw reads
2. Assemble the raw reads of the bacteriophage ϕ X174
3. Assess the quality of assembly
4. Explain any gap in knowledge of the bacteriophage via extraterrestrial intervention using the Sword in a Field principle: <https://www.youtube.com/watch?v=IWoyJrAAL-U>

Jan Jansen	s123456@student.dtu.dk	s123456
Fulan Al-Fulani	fulanf@aqua.dtu.dk	@fulanf
Maija Meikäläinen	s654321@student.dtu.dk	s654321
Zé Povinho	s456123@student.dtu.dk	s456123

----- END DO NOT EDIT -----

Group 1

Study: Integrating microbial and host transcriptomics to characterize asthma-associated microbial communities.

Question:

Steps:

1. Preprocessing:
 - a. Quality control (fastqc)
 - b. Trimming adapter tools, quality trimming and length filtering (bbduk)
2. Separating Human and Microbial reads (bbmap).
3. Taxonomical assignation of Microbial reads.
- 4.

Angeliki Kallia Spentza	s212881@student.dtu.dk	s212881
Javiera Laing Carmona	s196546@student.dtu.dk	s196546
Klara Alicja Kotkowska	s213028@dtu.dk	s213028
Mads Cort Nielsen	s120356@dtu.dk	s120356

Group 2

Study: <https://doi.org/10.12688/f1000research.55301.2>

Question: Differences in the psilocybin synthesis gene cluster of several Psilocybe genomes

Steps:

1. Preprocessing
 - a. Quality control (fastqc)
 - b. Trim adapter (fastp)
2. Align trimmed reads to the ref. genome of P. envy (bwa mem)
 - a. Coverage analysis
3. Variant analysis (bcftools or samtools)
4. Phylogenetic tree (treemix)

Agnete Karlslose	s163722@student.dtu.dk	s163722
Abril Diosdado Hernández	s196305@student.dtu.dk	s196305
Annette Lien	s202633@student.dtu.dk	s202633
Mikkel Eriksen	merik@dekom.dk	merik
Tram Bao Ngoc Nguyen	s216535@student.dtu.dk	s216535

Group 3

Study: Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality

<https://www.nature.com/articles/s43016-020-0129-3#Abs1>

Question: A comparison of four cheese types (blue cheese, cheddar, smoked cheese and tomme) and the differences in genus and phyla abundances, as well as similarities in genes.

Steps:

1. Preprocessing
 - a. Quality control (fastqc) - we did this before and after preprocessing
 - b. Trimming of adapters (bbduk)
 - c. Removing short and low quality sequences (sickle)
2. Taxonomic classification
 - a. Protein-based taxonomic classification (kaiju)
3. Gene analysis
 - a. De novo assembly (megahit)
 - b. Gene prediction (prodigal)
 - c. Gene catalog / clusters (mmseqs2)
 - d. Read-mapping (bwa + samtools)
 - e. Annotation (eggno-mapper)

Karolina Tudelska	s212846@student.dtu.dk	s212846
Caroline Jensen	s213427@student.dtu.dk	s213427
Maurice Kappelmeyer	s212710@student.dtu.dk	s212710
Jonas Dalsberg	s213551@student.dtu.dk	s213551
Patricia Manikova	s193712@student.dtu.dk	s193712

Group 4

Study: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3967933/?fbclid=IwAR0bNcQrpCs0cey8le0v4ZnvC0pFYGJP50Sec32WMtr-vFMOOZ1IJD6JaEU>

Question: Identify cancer specific gene fusions (CDKN2D-WDFY2) in RNA-Seq data and identify cancer specific gene expression.

Steps:

1. Quality control.
2. Map gene-expression data to reference genome using STAR.
3. Identify locations of mapped transcripts and compare with fusion genes
CDKN2D-WDFY2
4. Compare expression data from ovarian cancer cells to reference ovarian cell expression using Deseq2.
5. Identify short variants (SNPs and Indels) in RNAseq data according to GATK best practices...

Astrid Filt Beyer	s184248@student.dtu.dk	s184248
Caroline Uldahl Martensen	s183493@student.dtu.dk	s183493
Christian Peder Jacobsen	s165476@student.dtu.dk	s165476
Frederik Steensgaard Gade	fstga@dtu.dk	s184260
Søren Sandgaard	s173686@student.dtu.dk	s173686

Group 5

Study: RNA-seq analysis of bovine adipose tissue in heifers fed diets differing in energy and protein content

Question: Identify the impact of different diets (different protein and energy levels) on the transcription profile in adipose tissue in cattle.

Steps:

1. Quality Control of raw reads
2. Map gene expression using tophat2
3. Quantify gene expression using htSeq-count
4. Find differentially expressed genes using DESeq2
5. Find GO for the differentially expressed genes

Philip Martin Bjerre	s144068@student.dtu.dk	s144068
Christian Holm Johansen	s202770@student.dtu.dk	s202770
Rebecca Cecilie Grenov	s184344@student.dtu.dk	s184344
Mona Nourbakhsh	mnou@dtu.dk	mnou

Group 6

Study: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8715718/>

Question: Transcriptome analysis of Liraglutide effect on bone density in type 1 diabetic mice

Steps:

1. Quality check of raw reads
2. Index reference genome and align raw reads (HiSat2)
3. Sort (SAMtools) and count aligned reads (HtSeq-count)
4. Estimate differential expression of genes with DESeq2 (R)
5. Do GO enrichment analysis for differentially expressed genes

Paolo Federico	s212975@student.dtu.dk	s212975
Lea Jerman Plesec	s220067@student.dtu.dk	s220067
Oriol Portell I Vidal	s210419@student.dtu.dk	s210419
Oscar Solé I Salgado	s213691@student.dtu.dk	s213691
Sergio Esteban Echeverría	s212716@student.dtu.dk	s212716

Group 7

Study:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8228499/?fbclid=IwAR0lbubpUuxS3Qh0yEmXYM_Nr48Haa84Zsg8qJERC20yBekd2YSh76A4EMo

Question: Circulating cfDNA in the diagnosis of Cholangiocarcinoma.

- Steps:
- Fastp-Checking and trimming
 - BWA MEM-Reads aligned to reference genome
 - SAMtools-Sort the alignments
 - Picard Mark Duplicates-Remove PCR duplicates
 - VarScan2.4-Somatic variant calling
 - wANNOVAR-Functional annotation of somatic variants
 - Additional Filters
 - Statistical Analysis

Sara Húnfjörð Jósepsdóttir	s212952@student.dtu.dk	s212952
Aldís Helga Björgvinsdóttir	s212956@student.dtu.dk	s212956
Kolbrún Dóra Magnúsdóttir	s213732@student.dtu.dk	s213732
Adresya Suresh Azhakath	s212920@student.dtu.dk	s212920

Group 8:

Study: <https://www.biorxiv.org/content/10.1101/481242v1>

Question: Phylogeography of second plague pandemic by analysis of historical *Y. pestis* genomes

Steps:

1. Download raw reads from ENA (278 reads)
2. Download reference genomes (X out of 249 genomes+1 outgroup)

3. Preprocessing of de-multiplexed reads -removing adapters, read merging, filtering for sequencing quality and read lengths (AdapterRemoval)
4. Read mapping against CO92 genome, remove low quality mappings(BWA, SAMtools)
5. SNP calling using UnifiedGenotyper of the GATK
6. Identify all variant positions using MultiVCFAnalyzer, SnpEff to determine effect and annotation of each SNP
7. Perform Maximum likelihood phylogeny/neighbor joining using MEGA7 or RAxML (GTR model)
8. Select genomes with highest coverage for comparison
9. Remap selected genomes against CO92 genome
10. Calculate coverage across previously defined virulence genes (bedtools). Visualize heatmap using ggplot2 in R
11. Construct coverage plot for genomes against CO92 using Circos

Georgia Fanouraki	s202570@student.dtu.dk	s202570
Adikrishna Murali Mohan	s212940@student.dtu.dk	s212940
Siddhi Yash Jain	s212937@dtu.dk	s212937
Hossein Firoozabadi	hosfir@biosustain.dtu.dk	s216641

Group 9:

Study: <https://journals.asm.org/doi/10.1128/mSphere.00307-17>

Question: Reconstructing the bioinformatics pipeline for the HTS for the detection of virus contaminants in biotherapeutic products.

Steps:

1. Download RAW reads, Reference Genome, Virus Database (SRA)
2. FASTQC
3. Trimming
4. Indexing and aligning to the host genome (BWA)
5. Extract non-host reads (bam2fastq)
6. Extract model virus (spike-in) sequences from virus database
7. Align with the non-host reads against the database to create a consensus sequence (BWA)
8. Check for hits in GenBank to identify the specific isolate (BLAST)
9. Discussing the quality parameters of the bioinformatic pipeline
10. Skål! :)

Ranjitha Sekhar	s210271@dtu.dk	s210271
Varun Shankar	s212934@dtu.dk	s212934
Jake Pham	s213060@dtu.dk	s213060
Yu Liu	s212610@student.dtu.dk	s212610
Jesper Jensen	s974048@student.dtu.dk	s974048

Group 10:

Study: <https://www.nature.com/articles/s41564-018-0192-9>

Question: Does AMR in domesticated animals (pigs) differ between different countries?

Steps:

1. Preprocessing (Fastqc, adapter removal, Sickle, Host removal (bbMap))
2. Quality control (Nonpareil)
3. Taxonomic assignment with kaiju
4. Create map distribution of bacteria, classification table, statistics in R
5. AMR identification in ResFinder
6. Pathogen genome reconstruction with phyphlaN (time dependent)
7. Pick abundant species, download strain map strain variation (time dependent)

Tejs Nielsen	s220111@student.dtu.dk	s220111
Martina Vojtková	s212586@student.dtu.dk	s212586
andreas Kaas-mason	s213462@student.dtu.dk	s213462
Rikke Pedersen	s210417@student.dtu.dk	s210417
Carolina Rocha	s203014@student.dtu.dk	s203014

Group 11:

Study: <https://www.nature.com/articles/s41598-019-51562-6.pdf>

Question: Can WGS be used as a tool for characterizing epidemiological microorganisms like *M. tuberculosis* and assist epidemiological studies and disease control activities?

Steps:

1. Adapter removal, quality control
2. Aligning reads to reference genome
3. Exclude non-unique regions
4. Construct phylogenetic tree from SNP data
5. PCA data to see geographical clustering of variants

Emil-August Torp	s174341@student.dtu.dk	s174341
Yanan Zhao	s210319@student.dtu.dk	s210319
Ángela Miranda Segura	s210496@student.dtu.dk	s210496
Casper Rasmussen	s206220@student.dtu.dk	s206220
Huijiao Yang	s202360@student.dtu.dk	s202360

Group 12

Study: <https://link.springer.com/content/pdf/10.1186/s40169-016-0133-2.pdf>

Question: Do common antineoplastic drugs as trastuzumab and lapatinib produce cardiotoxicity?

Steps:

1. Quality control
2. Alignment to reference genome

3. Calculate read counts
4. Differential expression analysis
5. Enrichment analysis

Mathias Nørgaard Stokholm	s184257@dtu.dk	s184257
Inigo Miguelez Goyeneche	s205981@student.dtu.dk	s205981
Ziyu Zhang	s212657@dtu.dk	s212657
Danli Shen	s185387@student.dtu.dk	s185387
Sana Gull	s193271@student.dtu.dk	s193271