How to use Computerome 2

[**Computerome**](#_1u351a7ityge) **1**

[v KU-IT support](#_b39anj3cd0g9) 1

[v Login](#_bd52gfonka1a) 2

[v Projects, directory and file structure](#_ukjku68dxlie) 2

[v Running jobs](#_1bifb44xohxt) 3

[v Tools](#_ap8t2i77nxki) 4

[v Transfer data from/to Computerome](#_4zju1byh370b) 6

[v Check your usage](#_e6qj2gsqqe7) 6

[v Clean your space](#_8zpro6v0or8c) 7

[v Uploading data to GEO](#_3ekrtnqmvkyo) 7

[**Basic UNIX commands**](#_ecac7vrlbhtm) **8**

[v Files](#_b6uwb4tfat76) 8

[v Directories](#_1e6iq6dpif9c) 9

[v Useful tools](#_i9wmluu8zqj4) 9

[v Simple pipeline with |](#_d5yxp7ajvkjz) 9

[**Exchange data with internal or external collaborators**](#_2nxrn3ssbk7v) **10**

[**Online meetings**](#_44qhdzog53aw) **11**

# Computerome

[**https://www.computerome.dk/display/C2W/Computerome+2.0**](https://www.computerome.dk/display/C2W/Computerome+2.0)

(<https://www.computerome.dk/display/CW/Computerome+Wiki>)

## v KU-IT support

**I.**

Where to find **information** regarding C2:

DK: <https://kunet.ku.dk/medarbejderguide/Sider/It/Computerome-2.0.aspx>

EN: <https://kunet.ku.dk/employee-guide/Pages/IT/Computerome-2.0.aspx>

**II.**

How to send a **request** for having a new C2 **group** or for becoming a **user** of C2, herewith a free-of-charge “sandbox” user of C2 for 6 months:

Login to [IT self-service](http://sd.ku.dk/) with your KU username (such as “abc123@ku.dk”),

click on icon “New order” (1st ikon from left),

click on “Forsknings-IT / Research\_IT” under left-column Category, and

choose between one of the options.

**III.**

How to get **support and advises** on research-IT and data management:

Login to [IT self-service](http://sd.ku.dk/) with your KU username (such as “abc123@ku.dk”),

click on icon “New tickets and question” (2nd ikon from left),

click on “Forsknings-IT / Research\_IT” under left-column Category, and

choose between one of the options.

## **v Login**

1) Open a terminal

2) Login with ssh

**ssh username@ssh.computerome.dk**

3) Enter your password

4) Enter the PASSCODE you receive on your phone

More: <https://www.computerome.dk/display/C2W/SSH+login+to+Computerome+2.0>

## **v Projects, directory and file structure**

Predefined data structure:

1. /home/people/yourlogin => this is your private space. Only for configuration files. Limited to 10Gb. DO NOT STORE DATA THERE. Typically, you don’t need to use it.
2. /home/projects/yourproject => this is the shared space for your group/project. This is where you store data, results, etc.
3. /home/databases/ => this is a shared resource so that we don’t have to all download the same databases, e.g. genomes.

Inside each project, the following structure is defined:

**/home/projects/<project>:**

./apps

./apps/modulefiles

./backup

**./data** => typically here you have the group data (e.g. the processed genomics files)

**./people/<user>** => here you have your folder where you can do your things. Typically, I recommend creating a folder with the date each time you do something (e.g. 2020-08-27\_my-first-test)

./scratch

What project(s) do I have access to? You can type “id” to know:

E.g. for me (magmic)

*uid=234975(magmic) gid=234975(magmic) groups=234975(magmic),1217(danstem),1262(cu\_10077),1269(dan\_bri),1270(dan\_ser),1271(dan\_gra),1272(cpr\_cho),1273(cpr\_mai),1280(dan\_sem),1281(dan\_fer),1346(cpr\_gro),1362(cpr\_mon),1375(cu\_10129),1397(cpr\_gro\_bri),1423(cpr\_seq),1438(dan\_obe),1439(dan\_por),1440(cpr\_tay),1441(dan\_arn),1454(cpr\_ols),1466(dan\_and),1467(dan\_sed)*

## **v Running jobs**

!!! In general, and in the interest of the systems stability, jobs should **not** be run on the login node – it is **only** used as a **platform to submit** jobs into the high-performance cluster.

**1)** **First way is interactive job : iqsub**

**OK for relatively fast jobs, because unstable (relies on the connection between your computer and Computerome, can be broken if your computer enters sleep mode)**

· iqsub

· select your group

· select the time needed

· enter the number of processors needed

· wait for the node to be ready

· exit by typing “exit”

**2)** **Second way : shell script with qsub**

**Recommended for long jobs**

Create a file scriptname.pbs with the following content (to update based on your needs):

#!/bin/sh

### Note: No commands may be executed until after the #PBS lines

### Account information

#PBS -W group\_list=dan\_bri -A dan\_bri

### Job name (comment out the next line to get the name of the script used as the job name)

#PBS -N test

### Output files (comment out the next 2 lines to get the job name used instead)

#PBS -e test.err

#PBS -o test.log

### Send mail when job starts and finishes

#PBS -m abe

#PBS -M your\_email@sund.ku.dk

### Make output files readable by the group

#PBS -W umask=0027

### Number of nodes

#PBS -l nodes=1:ppn=8

### Memory

#PBS -l mem=80gb

### Requesting time - format is <days>:<hours>:<minutes>:<seconds> (here, 24 hours)

#PBS -l walltime=24:00:00

# Go to the directory from where the job was submitted (initial directory is $HOME)

echo Working directory is $PBS\_O\_WORKDIR

cd $PBS\_O\_WORKDIR

### Here follows the user commands:

# Define number of processors

NPROCS=`wc -l < $PBS\_NODEFILE`

echo This job has allocated $NPROCS nodes

# Load all required modules for the job

# This is where the work is done

echo Hello > test.file.txt

echo What a lovely day today >> test.file.txt

echo Congratulations, you have just created a lovely test file

Then you can submit this job by typing:

**qsub scriptname.pbs**

You can copy the file from /home/projects/cu\_10077/people/magmic/TRAINING/test.pbs

If you only have a few lines of code, you can leave it in the pbs file. If you have more, it is suggested to create a script.sh file with all the code and then to call the ./script.sh in your pbs file.

## **v Tools**

On computerome, you have to load the tools you will need.

To get a collection of ngs tools, you can do:

**module load ngs**

(Personally I don’t use that because I prefer to load each software independently to know exactly which software and which version I’m using).

Let’s say you want to use fastqc and don’t know how to load it.

You can search for a small part of the tool name

**module avail fastq**

And get something like that:

------------------------ /cm/local/.modulefiles\_cache/tools/modulefiles -------------------------

fastqc/0.11.2 fastqc/0.11.4 fastqc/0.11.5 fastqc/0.11.7 fastqtl/2.184 fastq-tools/0.8

Now you know which versions are available and can load the one you want (unless you want a specific one, chose the latest):

**module load fastqc/0.11.7**

If this tool has a dependency, you will get an error explaining that there is a dependency:

fastqc/0.11.7(39):ERROR:151: Module 'fastqc/0.11.7' depends on one of the module(s) 'perl/5.24.0 perl/5.20.2 perl/5.20.1'

fastqc/0.11.7(39):ERROR:102: Tcl command execution failed: prereq perl

You just need to add the dependency to your call:

**module load perl/5.24.0 fastqc/0.11.7**

You find there is another dependency:

fastqc/0.11.7(40):ERROR:151: Module 'fastqc/0.11.7' depends on one of the module(s) 'java/1.8.0-openjdk java/1.8.0 java/1.7.0-openjdk java/1.7.0'

fastqc/0.11.7(40):ERROR:102: Tcl command execution failed: prereq java

**module load perl/5.24.0 java/1.8.0-openjdk fastqc/0.11.7**

Success! Fastqc is now loaded and you can call it like that:

**fastqc --help**

Next time you just have to run that line for loading fastqc.

**module load perl/5.24.0 java/1.8.0-openjdk fastqc/0.11.7**

Some tools are based on python. In that case, you load anaconda first (e.g. cutadapt, bamCoverage, macs2).

**module load anaconda2/4.4.0**

Now the tool is loaded, e.g.:

**cutadapt --help**

If you don’t find your tool, just contact us and we’ll look together.

As most software, you can get more information and help on module by typing:

**module --help**

More: <https://www.computerome.dk/display/C2W/Installed+Software>

## v **Transfer data from/to Computerome**

1) Open a terminal (no connection to Computerome**):**

2) From my computer to Computerome :

scp /Users/path/to/file username@ssh.computerome.dk:/home/path/to/the/destination/

3) From Computerome to my computer :

scp username@ssh.computerome.dk:/home/path/to/file/ /Users/path/to/destination/

Note :

- if you want to copy one directory, use scp –r

- if you want to copy all the files

§ terminated by “.bam” for example, use \*.bam

§ started with “Sample1” for example, use Sample1\*

## **v Check your usage**

At the end of every month, have a look at your usage to get an idea of how much compute time you have used.

You first need to load the required script:

**module load usage\_script/2.0**

> your overall usage since you started using your account

usage

> your usage in a given year

usage -y 2020

The important number here is the total CPH hours. Say you have 5000.

Given that the typical price is DKK 0.13/CPU core hour, you get an idea of the “price” of what you have computed (here 5000 x 0,13 = 650 DKK). As long as you use the group id to submit your jobs, this should not be billed to you or your group (it is paid by the platform for now), but it is useful to know the value of the computations.

If this number becomes higher than 2000, it might be good that you come to us to discuss what large computations you are doing.

If this number becomes higher than 5000, please come to us.

## **v Clean your space**

Here are some simple cleaning rules

* All fastq files are zipped (.fastq.gz): gzip \*.fastq
* When the reads have been demultiplexed into several fastq.gz files, delete the unassigned reads: rm Undetermined\_S0\_R\*\_001.fastq.gz
* If 2 versions (zipped and unzipped) of a file exists, the unzipped is deleted
* Delete intermediate files that can be re-generated from the fastq.gz files, for instance:
  + If you have the BAM files, delete the SAM files used to obtain the BAM
  + If you have unsorted and sorted BAM, keep only the sorted ones
  + If you have filtered BAM files, remove the full one you had before filtering

## **v Uploading data to GEO**

If you have non-human sequencing data, you may want to upload them to GEO (or other repositories).

Documentation on **Submitting high-throughput sequence data to GEO**

<https://www.ncbi.nlm.nih.gov/geo/info/seq.html>

You will need to upload your data and then notify GEO. If your data are larger than 4TB, they recommend that you contact GEO before uploading.

If your data are on Computerome, you can upload directly from Computerome using for instance ncftp

In short:

1. Create a folder on Computerome with your GEO user name (e.g. geouser)
2. Put all your submission files in there (if fastq.gz files from the data folder, please do not move the original files, make a copy with cp and after the submission is finished, you can remove your submission folder).
3. Start an iqsub session (with at least 8h so the upload does not get interrupted) (for iqsub, see 1) above)
4. From outside this geouser folder, run:  
   module load ncftp/3.2.5  
   ncftp
5. You are now inside the ncftp tool and can run the following:  
   set passive on

set so-bufsize 33554432

open ftp://geo:33%259uyj\_fCh%3FM16H@ftp-private.ncbi.nlm.nih.gov

**Double check the copy-paste: check that it is exactly the same and that no funny changes happened with % or something else.. If unsure, paste in a text only file or type it directly in the command line**

1. Finally, you launch the transfer (using your folder name instead of geouser):  
   put -R geouser

More details here if needed: <https://www.ncbi.nlm.nih.gov/geo/info/submissionftp.html>

# Basic UNIX commands

These are commands that you can use when you are connected to Computerome.

## **v Files**

* **ls** --- lists your files   
   **ls -l** --- lists your files in 'long format', which contains lots of useful information, e.g. the exact size of the file, who owns the file and who has the right to look at it, and when it was last modified.
* **more *filename*** --- shows the first part of a file, just as much as will fit on one screen. Just hit the space bar to see more or **q** to quit. You can use **/*pattern*** to search for a pattern.
* **mv *filename1 filename2*** --- moves a file (i.e. gives it a different name, or moves it into a different directory (see below)
* **cp *filename1 filename2*** --- copies a file
* **rm *filename*** --- removes a file. It is wise to use the option rm -i, which will ask you for confirmation before actually deleting anything.
* **diff *filename1 filename2*** --- compares files, and shows where they differ
* **wc -l *filename*** --- tells you how many lines there are in a file

## **v Directories**

* **mkdir *dirname*** --- make a new directory
* **cd *dirname*** --- change directory. You basically 'go' to another directory, and you will see the files in that directory when you do 'ls'. You always start out in your 'home directory', and you can get back there by typing 'cd' without arguments. 'cd ..' will get you one level up from your current position.
* **pwd** --- tells you where you currently are.

Note: Use the tab key to auto complete the names of directories and files

## **v Useful tools**

* **vi *filename*** --- is an editor that lets you create and edit a file.

1) **i** --- for insertion

2) escape key

3) **:wq!** --- writes and quits

or **:q!** --- quits without saving

* **grep/egrep** commandis to search for a given string in a single file

**grep** “chrM” ***filename*** --- only prints lines containing “chrM”

**grep -v** “chrM” ***filename*** --- only prints lines **not** containing “chrM”

**egrep -v** "^@" ***filename*** --- do not print lines **starting** with “@”

## **v Simple pipeline with |**

Pipe = | (alt + i )

The standard shell syntax for pipelines is to list multiple commands, separated by vertical bars | ("pipes" in common Unix verbiage)

Ex:

egrep -v "^@"filename.sam | wc -l

--- counts the lines of a .sam file without the header

(bowtie output is in .sam format)

samtools view filename.bam | more

--- show the first lines of a .bam file

(.bam is binary, the samtools view command is required to convert it)

# Exchange data with internal or external collaborators

We often need to exchange data with either internal or external collaborators, but the information on how to accomplish this is scattered.

Below are given three options for sharing data:

1. Email attachment (small data set, limit ~<10-20Mb)

2. Bluewhale (medium, <4Gb)

3. FileSender (large, <256Gb)

------------------------------------------------------

2.

Bluewhale (Dropbox-like)

Website: [bluewhale.ku.dk](http://bluewhale.ku.dk/)

Features:

Share files with others (<4Gb)

Login with KU-credentials to upload - only KU employees can upload

Everybody can download - including external collaborators

Also available as plugin for email client

Encryption option

Documentation: See KU-IT Bluewhale guide

------------------------------------------------------

3.

FileSender (Dropbox-like)

Website: <https://filesender.deic.dk>

Features:

Share files with others (<256Gb)

Login with KU-credentials (Choose 'University of Copenhagen' in list of login provider)

KU employees can give upload-voucher to external collaborators i.e. non-KU person can upload data for you to download!

Allows individual file sizes >2Gb, if browser supports html5

Encryption option

Documentation: <https://www.deic.dk/en/node/558>

------------------------------------------------------

# Online meetings

Some information about access to professional online meeting tools.

UCPH staff members have free access to **Zoom** and **Adobe connect** via the Danish e-infrastructure cooperation [**deic.**](https://www.deic.dk/)

For **Zoom** do the following.

- Go to: <https://deic.zoom.us/>

- Choose host a meeting

- Choose University of Copenhagen (Københavns Universitet)

- Login with your KU credentials

- Create meeting time.

- You will get a link to give to your guests.

Zoom also allows everybody to share their screen and the administrator can see who is participating and receive questions or “hand-raises” and manage the meeting.

**Adobe connect** can be reached under:

[https://c.deic.dk](https://c.deic.dk/)

Then follow similar steps as above.