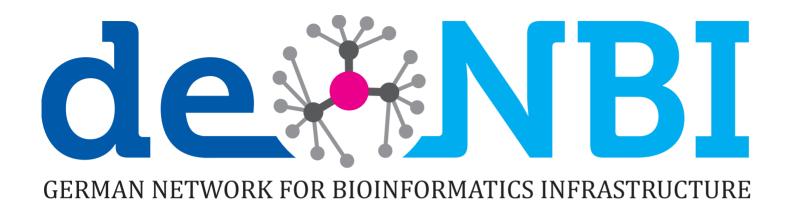
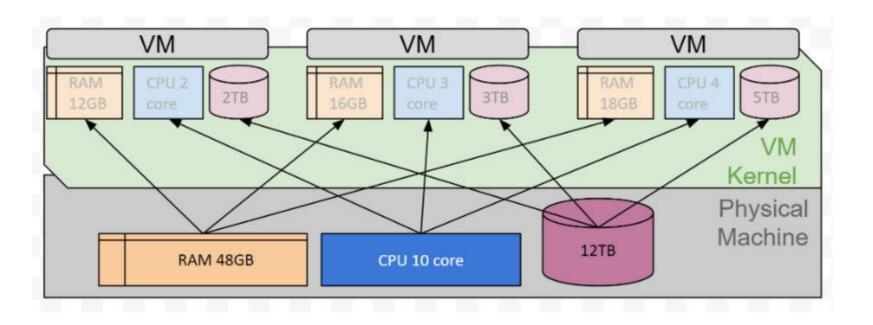
Functional microbiome research – bioinformatics section

Day 1 – intro to linux

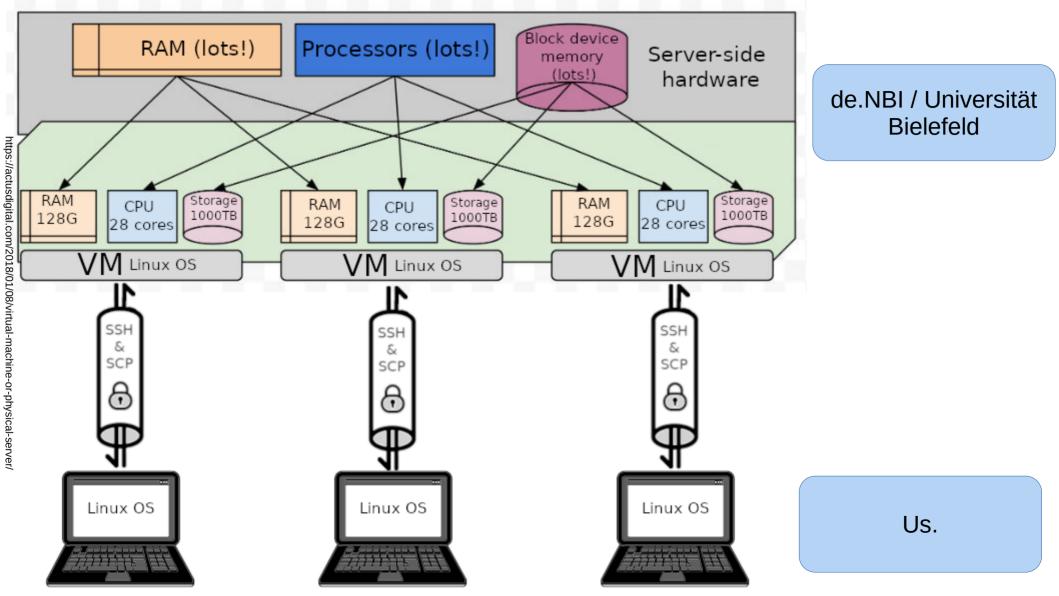


First, thanks to de.NBI. They are amazing.

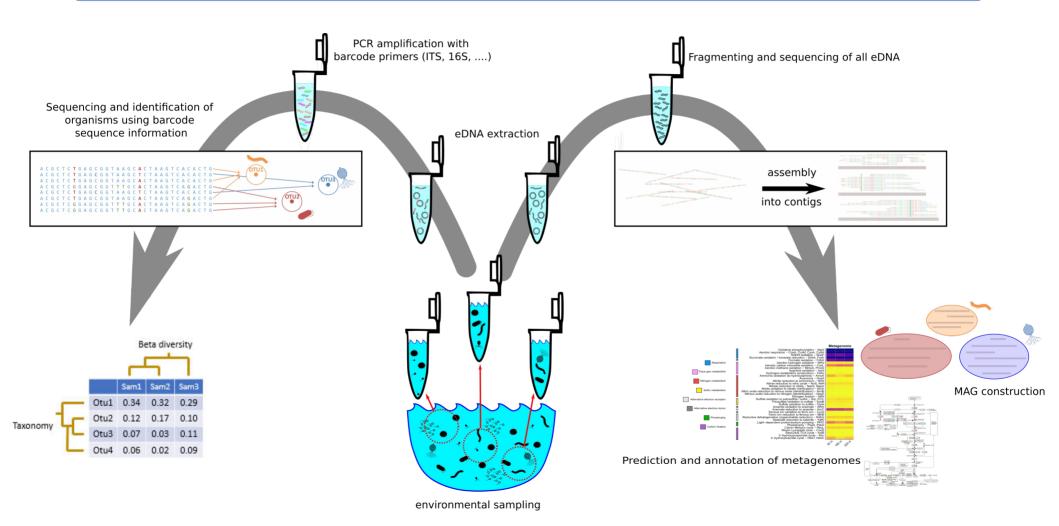
Distributed computing and compute clusters are essential tools for scientists.



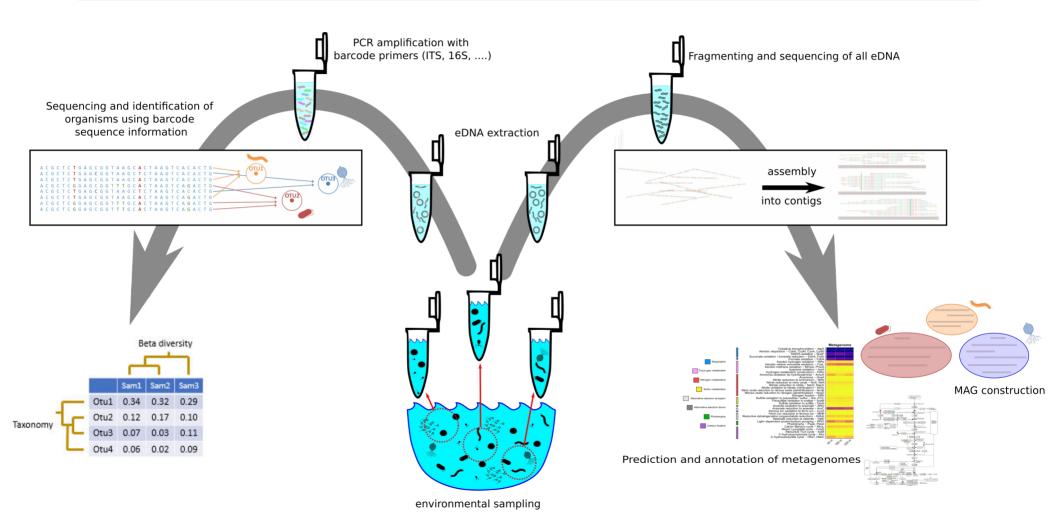
We use our local computers to talk to shared remote servers that have lots of high-powered processors, and storage and RAM.



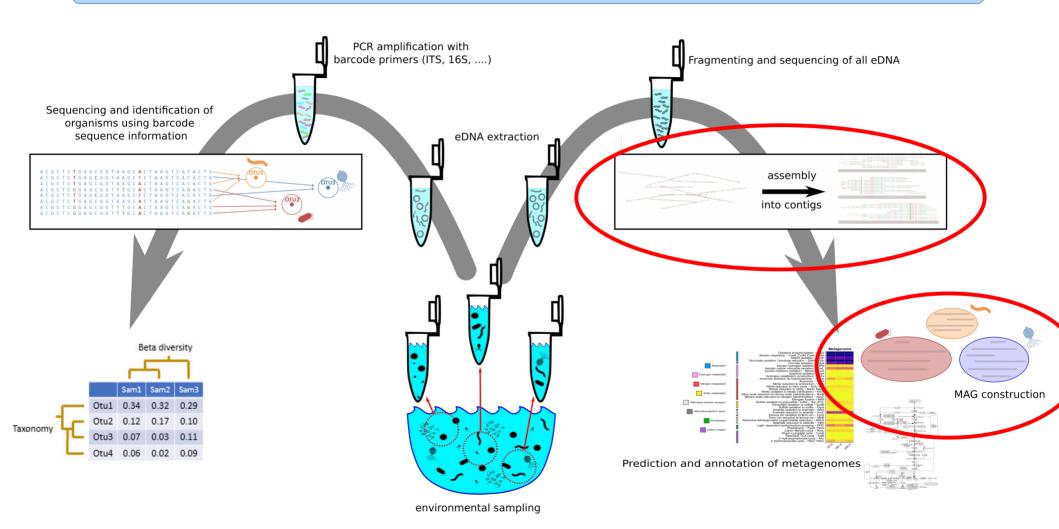
Review – what do we do with DNA sequence data?



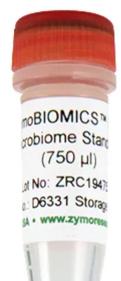
Why do we need distributed computing to do this?



Because some steps take ~100 gig of RAM (or more!) and lots of processors.



Dataset 1: ZymoBIOMICS HMW DNA Standard.



	Theoretical Composition (%)				
Species	Genomic DNA	16S Only ¹	16S & 18S¹	Genome Copy ²	Cell Number ³
Pseudomonas aeruginosa	14	5.1	4.6	7.8	7.9
Escherichia coli	14	12.4	11.2	10.9	11.0
Salmonella enterica	14	12.7	11.4	11.2	11.2
Enterococcus faecalis	14	12.1	10.9	18.8	18.8
Staphylococcus aureus	14	19	17.1	19.6	19.6
Listeria monocytogenes	14	17.3	15.6	17.8	17.9
Bacillus subtilis	14	21.4	19.2	13.2	13.2
Saccharomyces cerevisiae	2	NA	10	0.63	0.32

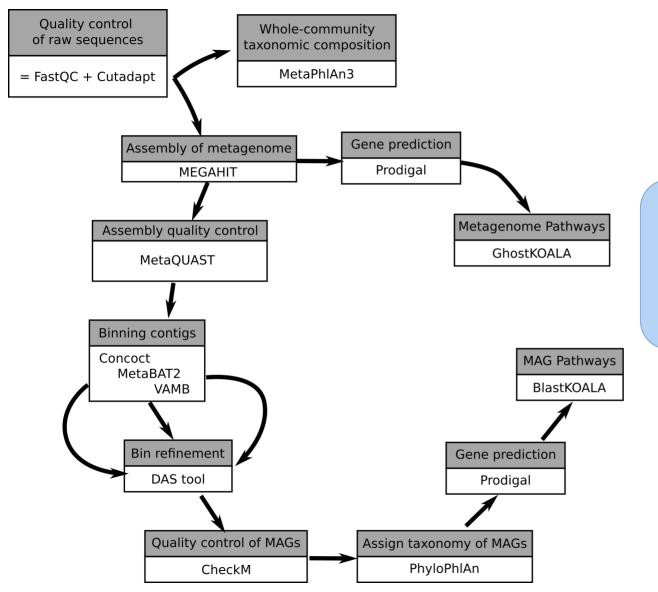
OPEN

Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing

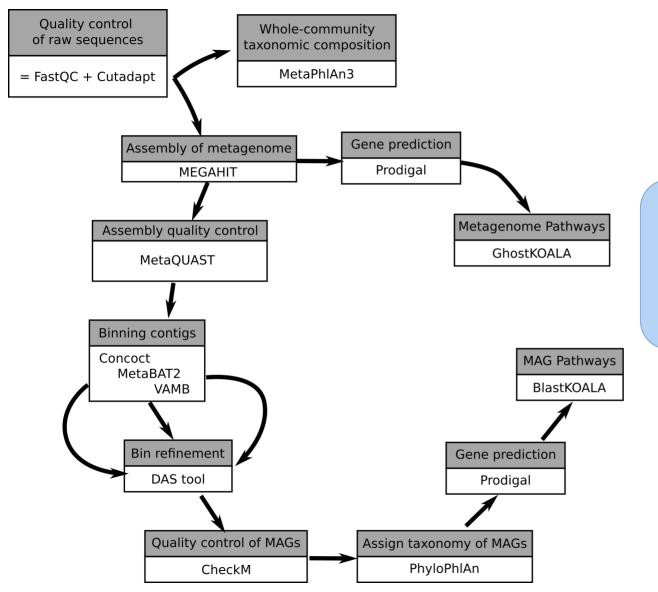
Mantas Sereika [©] ^{1,4}, Rasmus Hansen Kirkegaard [©] ^{1,2,4}, Søren Michael Karst [©] ¹, Thomas Yssing Michaelsen ¹, Emil Aarre Sørensen [©] ¹, Rasmus Dam Wollenberg ³ and Mads Albertsen [©] ¹ [∞]



Fredericia – Centralrenseanlægget is the second largest wastewater treatment plant in Denmark



We will use various scientific software packages to process our sequence data into a metagenome and Metagenome-Assembled-Genomes (MAGs).



We will do these steps with both illumina and nanopore data, for our learning process with the Zymo Mock Community data.

pwd – print the current working directory

Is – list the contents of the current directory

cd – change directory

cp – copy a file

mv - move or rename a file

mkdir – make a new directory

echo - print something out to the display

cat – concatenate two or more files

rm [-r] – remove a file. Essentially deletes a file forever. Be careful!!

sudo – execute another command that requires root privileges

top – opens up a real time display that shows you the "busiest" processes ongoing in your machine. After you start the program, press i to see only active processes, press e to make memory units more readable, and press 1 then t to keep a close eye on your cores.

man – print the reference manual for a utility, command, or program, if available. If not, try
the --help flag for any given program without a manual.

head – show top several lines from a text file

tail - show last several lines from a text file

less - interactively open a text file for reading

Special symbols are very important to the linux command line.

- home directory (try with cd)
- current directory (try with cd). Also filenames that start with "." are "hidden".
- .. parent directory (try with cd)
- > direct the output of a process to a file
- \$ variable expansion
- pipe, connects the output from one command to another
- ; command separator
- / directory separator (compare to Windows!)
- \ escape a character, so that it is no longer a "special character"
- & send a process to run in the background
- = variable assignment
- * wildcard for multiple characters (try with ls)
- ? wildcard for single character (try with ls)
- # comment (place it before a command and see what happens)

top – opens up a real time display that shows you the "busiest" processes ongoing in your machine. After you start the program, press i to see only active processes, press e to make memory units more readable, and press 1 then t to keep a close eye on your cores.

man – print the reference manual for a utility, command, or program, if available. If not, try
the --help flag for any given program without a manual.

head – show top several lines from a text file

tail - show last several lines from a text file

less - interactively open a text file for reading

Special command: wget

We'll keep track of what we do in scripts that will be kept in an online github repository. Let's practice grabbing the latest version:

wget https://raw.githubusercontent.com/danchurch/FunctionalMicrobiomePractical2022/main/funmic2023/funBASHterminalScript.txt

(Maybe better to follow the link in the skript PDF on page 20, or download this presentation, then copy-paste)

First task: connect to VMs Milan ssh -p 30284 -i /path/to/your/privateKey ubuntu@129.70.51.6

Julia ssh -p 30381 -i /path/to/your/privateKey ubuntu@129.70.51.6

Maria ssh -p 30355 -i /path/to/your/privateKey ubuntu@129.70.51.6

Anja ssn -p 30355 -i /patn/to/your/privatekey ubuntu@129.70.51.6

ssh -p 30452 -i /path/to/your/privatekey ubuntu@129.70.51.6

Anja ssh -p 30452 -i /path/to/your/privateKey ubuntu@129.70.51.6

Susanne ssh -p 30461 -i /path/to/your/privateKey ubuntu@129.70.51.6

Rebekka ssh -p 30372 -i /path/to/your/privateKey ubuntu@129.70.51.6

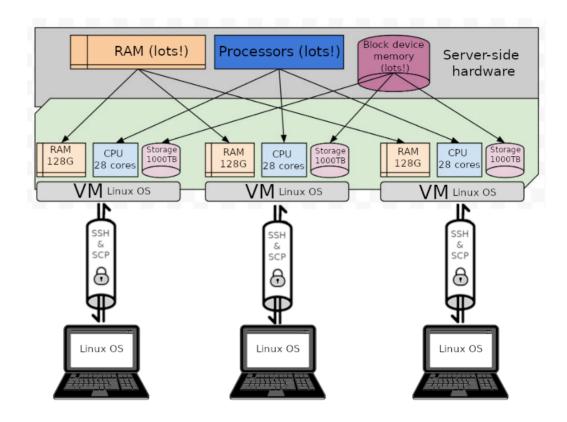
key=/home/daniel/.ssh/funmic2023.pub ssh -p 30500 -i \$key ubuntu@129.70.51.6

Or try with a

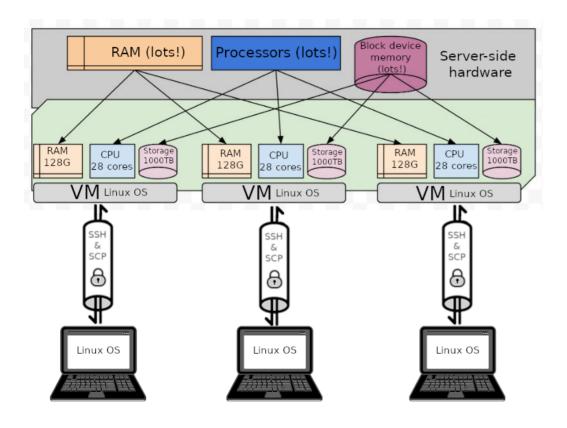
shell variable:

First task: connect to VMs

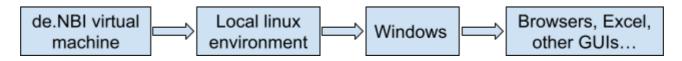
cd /vol/danBot

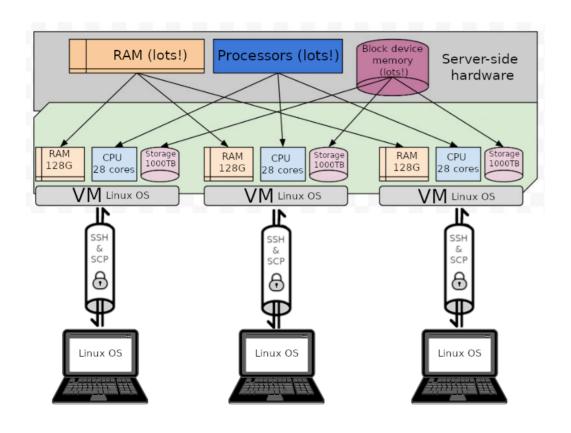


Once you are in, try to find your storage volume.

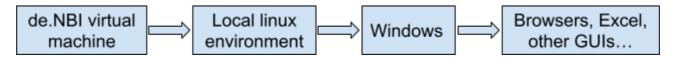


You will often need to get files from your de.NBM VM, to inspect them locally either in your linux or windows evironment.



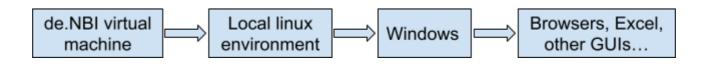


We will use the **scp** program for this.



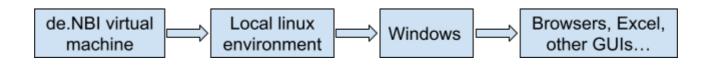
```
fileOnVM="/vol/danBot/helloWorld.txt"
mySSHkey="/home/daniel/.ssh/funmic2023.pub"
scp -i /home/daniel/.ssh/funmic2023 -P 30500 ubuntu@129.70.51.6:$fileOnVM .
```

Try **scp**, to get "helloWorld.txt" from your de.NBI VM to your local machine.

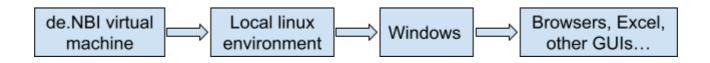


```
fileOnVM="/vol/danBot/helloWorld.txt"
mySSHkey="/home/daniel/.ssh/funmic2023.pub"
scp -i /home/daniel/.ssh/funmic2023 -P 30500 ubuntu@129.70.51.6:$fileOnVM .
```

(Can copy and paste from scripts, but make sure you understand!)



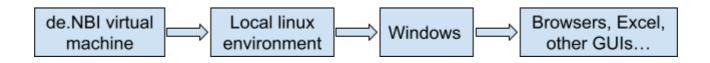
Once you have your file in your local linux environment, you will probably want to look at it in your windows environment, too.

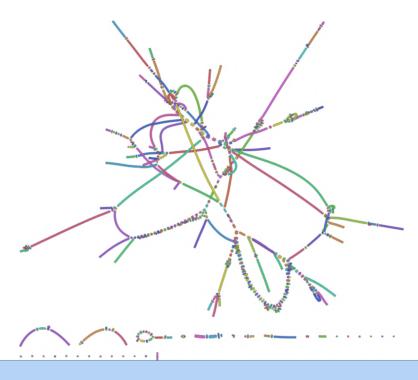


If you are using WSL in windows, the try placing the file here:

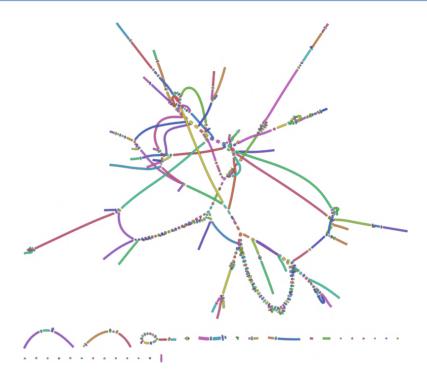
cd /mnt/c/

See if you can find it in windows by looking at your C:\ directory. Considering making a special folder for this kind of file transfer.

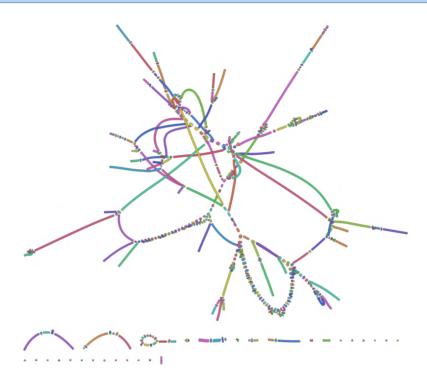




Assembly of metagenomes is one of the most computationally intensive tasks we will ask our de.NBI machines to do for us. We will review metagenome assembly a bit romorrow, but let's let's get our first assembly started now, so it can as long as it needs to.



This will be the first of two assemblies we do, since we have both an illumina and a nanopore dataset from the same wastewater sample. We will start with the nanopore data, since it takes much more time and computing resources.



This also gives a chance to see a complex command in action, that requires us to use much of what we just learned about the linux command line.

From my de.NBI machine, the command looks like this:

```
conda activate flye
mkdir -p /vol/danBot/assemblies/sludge/nanopore/
reads=/vol/danBot/datasets/zymoMC/nanopore/ERR7287988 shortened.fastq
outdir=/vol/danBot/assemblies/sludge/nanopore/
nohup flye --meta \
   --threads 25 \
   --out-dir Soutdir \
    --nano-hq $reads &
```

(You can see the full help file by typing flye --help)

Did that work?

```
daniel@daniel-ThinkPad-X1-Yoga-3rd: ~/Documents/teaching/functionalMicrobiomes/FunctionalMicrobiomePractical2022_repo/funmic2023
Comp. 10:26:51 up 4 days, 1:03, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1:03, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1:03, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1:03, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1:03, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1:03, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1:03, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1 user, load average: 0.00, 0.04, 0.02

Top. 10:26:51 up 4 days, 1 user, load average: 0.00, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.0
                 NB Mem : 128827.5 total, 124423.4 free, 381.7 used, 4022.4 buff/cache
NB Swap: 1024.0 total, 1024.0 free, 0.0 used. 127306.2 avail Mem
                 PID USER PR NI VIRT RES SHR S %CPU %MEM TIME+ COM 60564 ubuntu 20 0 11332 3964 3188 R 0.3 0.0 0:00.09 top
```

Let's check with the top program.

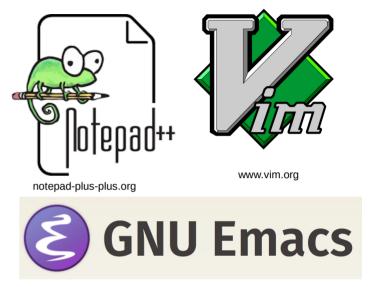
```
pinMode(LED, OUTPUT):
  oid loop()
   digitalWrite(LED, HIGH);
  digitalWrite(LED, LOW);
delay(250);
                                                                                                                                    avr-g++ -o .<mark>pioenvs/uno/firmware.elf</mark> -Os -mmcu=atmega328p -Wl,--gc-sections,--relax .pio
envs/uno/src/blink.o -L/Users/zach/.platformio/packages/ldscripts -L.pioenvs/uno -Wl,--s
                                                                                                                                   tart-group -lm .pioenvs/uno/libFrameworkArduinoVariant.a .pioenvs/uno/libFrameworkArdu
                                                                                                                                   avr-objcopy -0 ihex -R .eeprom .pioenvs/uno/firmware.elf .pioenvs/uno/firmware.hex
"avr-size" —mcu=atmega328p -C -d .pioenvs/uno/firmware.elf
                                                                                                                                   Program: 998 bytes (3.0% Full)
(.text + .data + .bootloader)
                                                                                                                                   Data: 9 bytes (0.4% Full)
(.data + .bss + .noinit)
                                                                                                                                  Built target platformio_build
                                                                                                                                  Compilation finished at Fri Dec 25 23:08:44
13:12 U -[blink]src/blink.cpp
```

With scientific computing, it can sometimes take hours to build, debug, and successfully execute a single program, sometimes just one line of code.

```
pinMode(LED, OUTPUT):
  oid loop()
   digitalWrite(LED, HIGH);
  digitalWrite(LED, LOW);
delay(250);
                                                                                                                                   avr-g++ -o .<mark>pioenvs/uno/firmware.elf</mark> -Os -mmcu=atmega328p -Wl,--gc-sections,--relax .pio
envs/uno/src/blink.o -L/Users/zach/.platformio/packages/ldscripts -L.pioenvs/uno -Wl,--s
                                                                                                                                   tart-group -lm .pioenvs/uno/libFrameworkArduinoVariant.a .pioenvs/uno/libFrameworkArdu
                                                                                                                                   avr-objcopy -0 ihex -R .eeprom .pioenvs/uno/firmware.elf .pioenvs/uno/firmware.hex
"avr-size" —mcu=atmega328p -C -d .pioenvs/uno/firmware.elf
                                                                                                                                   Program: 998 bytes (3.0% Full)
(.text + .data + .bootloader)
                                                                                                                                   Data: 9 bytes (0.4% Full)
(.data + .bss + .noinit)
                                                                                                                                  Built target platformio_build
                                                                                                                                  Compilation finished at Fri Dec 25 23:08:44
13:12 U -[blink]src/blink.cpp
```

You will often want to return to your commands days or months later. Other scientists will want to know precisely what commands and settings you used.

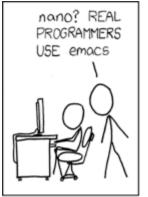
```
pinMode(LED. OUTPUT):
  oid loop()
  digitalWrite(LED, LOW);
delay(250);
                                                                                                                 vs/uno/src/blink.o -L/Users/zach/.platformio/packages/ldscripts -L.pioenvs/uno -Wl,
                                                                                                               avr-objcopy -0 ihex -R .eeprom .pioenvs/uno/firmware.elf .pioenvs/uno/firmware.he
"avr-size" ---mcu=atmega328p -C -d .pioenvs/uno/firmware.elf
                                                                                                                          998 bytes (3.0% Full)
                                                                                                                              9 bytes (0.4% Full)
                                                                                                             Built target platformio_build
                                                                                                              Compilation finished at Fri Dec 25 23:08:44
13:12 U -[blink]src/blink.cpp
```



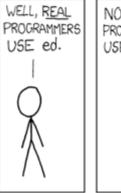
www.gnu.org/software/emacs/tour/

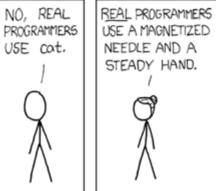
Keep a running text-file script of all that you do. Comment it with your own mental notes so you can remember what you did and why in human language.

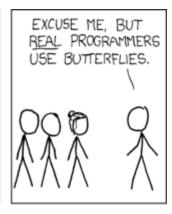
Pick a good **text editor** (not word processor!!) and learn to use it.



















THESE CAUSE MOMENTARY POCKETS OF HIGHER-PRESSURE AIR TO FORM,

WHICH ACT AS LENSES THAT DEFLECT INCOMING COSMIC RAYS, FOCUSING THEM TO STRIKE THE DRIVE PLATTER AND FLIP THE DESIRED BIT.

