

**Lab Book**  
*Cientific Initiation - Coral Metagenomes*  
Letcia Costa Cavalcante

Pedro Meirelles  
Institute of Biology - UFBA

2018

# Sumário

<b>1</b>	<b>May 2018</b>	<b>2</b>
1.1	13 . . . . .	2
1.1.1	Learning L <sup>A</sup> T <sub>E</sub> X . . . . .	2
1.1.2	Math environment . . . . .	3
1.1.3	15 - Short-term project proposal . . . . .	4
<b>2</b>	<b>Creation of data base of metagenomes and genomes</b>	<b>5</b>
2.1	28 . . . . .	5
2.1.1	Bibliographic search for genomes . . . . .	5
2.2	28 . . . . .	9
2.2.1	Bibliographic search for metagenomes . . . . .	9
<b>3</b>	<b>Download of metagenomes</b>	<b>12</b>
3.1	Download of mg-rast files . . . . .	12
3.2	Download of NCBI metagenomes . . . . .	12
<b>4</b>	<b>Format Conversion of NCBI metagenomes</b>	<b>14</b>
<b>5</b>	<b>Quality filter</b>	<b>15</b>
<b>6</b>	<b>Uniformity filter (size and N bases)</b>	<b>16</b>
6.1	Command line . . . . .	16
<b>7</b>	<b>Profiling metagenomes</b>	<b>18</b>
7.1	Mg-Rast metagenomes . . . . .	18
7.2	Kraken-biom . . . . .	18
7.3	Teste com o kraken no scratch . . . . .	19
<b>8</b>	<b>Functional annotation of metagenomes</b>	<b>21</b>
<b>9</b>	<b>references</b>	<b>23</b>
<b>10</b>	<b>Softwares, instalacao e linhas</b>	<b>25</b>
10.1	Profiling metagenomes . . . . .	25

# Capítulo 1

## May 2018

### 1.1 13

#### 1.1.1 Learning L<sup>A</sup>T<sub>E</sub>X

- Working folder: *path*

L<sup>A</sup>T<sub>E</sub>X is a high-quality typesetting system, available as free software, which allows to produce scientific or technical documents [?]. I am using L<sup>A</sup>T<sub>E</sub>X to create a Bioinformatics Lab Book. To compile my Lab Book, I can use command lines ([pdflatex](#) and [bibtex](#)). Afterwards I can visualise the produced *.pdf* file with evince or another reader. Alternatively, I can use a Latex editor, such as TexWorks (<https://www.tug.org/texworks/>), which allows me to write the code and control the *pdf* file in the same environment (Figure 8.1).

To compile the *.tex* file in the command line:

```
$pdflatex lab-book
$bibtex lab-book
$pdflatex lab-book
$pdflatex lab-book
```

To visualise the *.pdf*:

```
$evince lab-book.pdf &
```

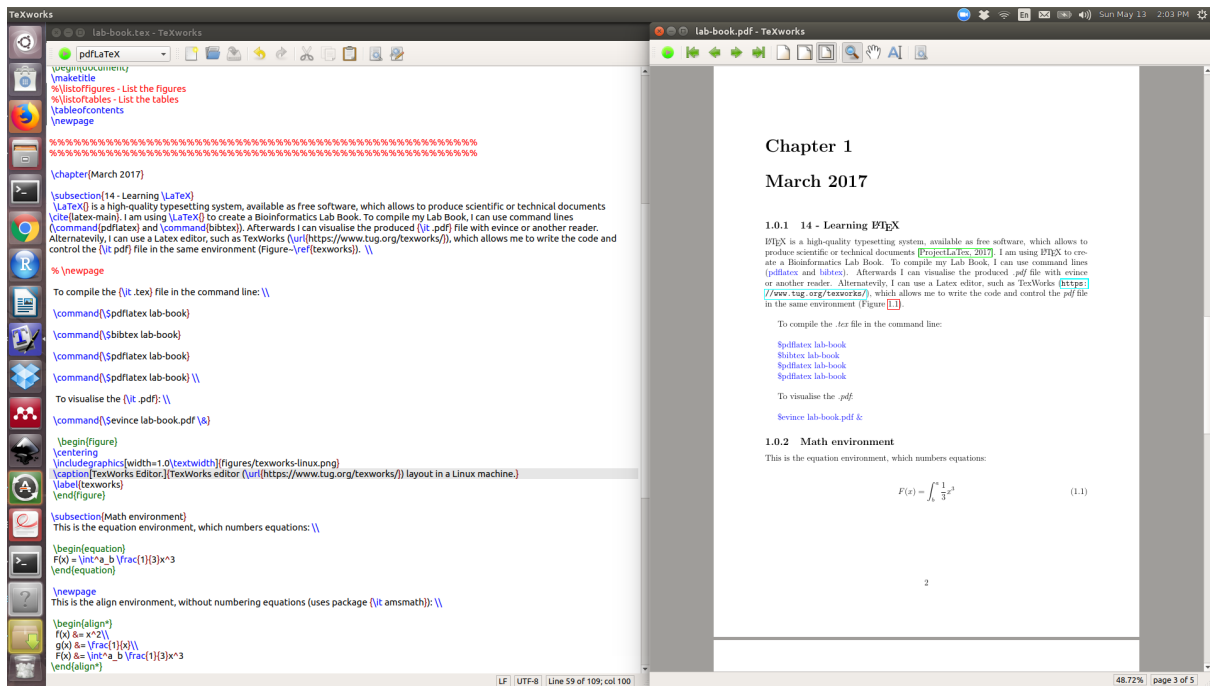


Figura 1.1: TexWorks editor (<https://www.tug.org/texworks/>) layout in a Linux machine.

## 1.1.2 Math environment

This is the equation environment, which numbers equations:

$$F(x) = \int_b^a \frac{1}{3} x^3 \quad (1.1)$$

This is the align environment, without numbering equations (uses package *amsmath*):

$$\begin{aligned} f(x) &= x^2 \\ g(x) &= \frac{1}{x} \\ F(x) &= \int_b^a \frac{1}{3} x^3 \end{aligned}$$

### 1.1.3 15 - Short-term project proposal

Some text here. Including and referencing a table (table 1.1).

- First numbered list item
- Second numbered list item

Tabela 1.1: table0

species	changes	score
Macaque	4	0.0
Human	2	14.9
Orangutan	0	0.0
Pan	0	0.0
Gorilla	0	0.0

# Capítulo 2

## Creation of data base of metagenomes and genomes

### 2.1 28

#### 2.1.1 Bibliographic search for genomes

Found a new possibility of phyla list. Because of this, there are four possibilities of list of microorganisms phyla, one of them, the SILVA database, is based in RNA sequences:

- The list of Prokariotic names with stading nomenclature <http://www.bacterio.net/-classifphyla.html>
- SILVA database LSU(large subunit of ribosome) <https://www.arb-silva.de/browser/lsu/>
- SILVA database SSU(small subunit of ribosome) <https://www.arb-silva.de/browser/ssu/>
- PATRIC GENOMES [https://www.patricbrc.org/view/Taxonomy/2#view\\_tab=taxontree](https://www.patricbrc.org/view/Taxonomy/2#view_tab=taxontree)

The list of articles used until now is:

- 10.1038/nature14486
- 10.1038/ismej.2013.111
- 10.1038/ismej.2013.174
- 10.1038/ismej.2016.43
- 10.1038/nature12352
- 10.1038/nature14486
- 10.1038/nature21031
- 10.1038/ismej.2015.233

- 10.1038/ncomms13219
- 10.1073/pnas.0801980105
- 10.1111/1462-2920.13362
- 10.1126/science.1132690
- 10.1186/s40168-015-0077-6

The list os correspondent phyla and articles is above

Tabela 2.1: table 1

DOI	Phylum
10.1038/nature14486	Candidatus Falkowbacteria
10.1038/nature14486	Candidatus Kuenenbacteria
10.1038/nature14486	Candidatus Magasanikbacteria
10.1038/nature14486	Candidatus Uhrbacteria
10.1038/nature14486	Candidatus Moranbacteria
10.1038/nature14486	Candidatus Azambacteria
10.1038/nature14486	Candidatus Yanofskybacteria
10.1038/nature14486	Candidatus Jorgensenbacteria
10.1038/nature14486	Candidatus Wolfebacteria
10.1038/nature14486	Candidatus Giovannonibacteria
10.1038/nature14486	Candidatus Nomurabacteria
10.1038/nature14486	Candidatus Campbellbacteria
10.1038/nature14486	Candidatus Adlerbacteria
10.1038/nature14486	Candidatus Kaiserbacteria
10.1038/nature14486	C. S. yataiensis
10.1038/nature14486	Pacebacteria
10.1038/nature14486	Candidatus Collierbacteria
10.1038/nature14486	Candidatus Beckwithbacteria
10.1038/nature14486	Candidatus Roizmanbacteria
10.1038/nature14486	Candidatus Saphirobacteria
10.1038/nature14486	Candidatus Amesbacteria
10.1038/nature14486	Candidatus Woesebacteria
10.1038/nature14486	Candidatus Gottesmanbacteria
10.1038/nature14486	Candidatus Levybacteria
10.1038/nature14486	Candidatus Daviesbacteria
10.1038/nature14486	Candidatus Curtissbacteria
10.1038/nature14486	WWE3
10.1038/nature14486	CPR3
10.1038/nature14486	WS6
10.1038/nature14486	Candidatus Berkelbacteria
10.1038/nature14486	Candidatus Peregrinibacteria
10.1038/nature14486	Candidatus Gracilibacteria
10.1038/nature14486	CPR2
10.1038/nature14486	Kazan
10.1038/nature14486	Saccharibacteria (TM7)
10.1038/nature14486	SR1
10.1038/ncomms13219	Candidatus Kerfeldbacteria
10.1038/ncomms13219	Candidatus Komeilibacteria
10.1038/ncomms13219	Candidatus Andersenbacteria
10.1038/ncomms13219	Candidatus Ryanbacteria
10.1038/ncomms13219	Candidatus Niyogibacteria



10.1038/ncomms13219	Candidatus Tagabacteria
10.1038/ncomms13219	Candidatus Terrybacteria
10.1038/ncomms13219	Candidatus Vogelbacteria
10.1038/ncomms13219	Candidatus Zambryskibacteria
10.1038/ncomms13219	Candidatus Taylorbacteria
10.1038/ncomms13219	Candidatus Sungbacteria
10.1038/ncomms13219	Candidatus Brennerbacteria
10.1038/ncomms13219	Candidatus Spechtbacteria
10.1038/ncomms13219	Candidatus Staskawiczbacteria
10.1038/ncomms13219	Candidatus Wildermuthbacteria
10.1038/ncomms13219	Candidatus Portnoybacteria
10.1038/ncomms13219	Candidatus Woykebacteria
10.1038/ncomms13219	Candidatus Blackburnbacteria
10.1038/ncomms13219	Candidatus Chisholmbacteria
10.1038/ncomms13219	Candidatus Buchananbacteria
10.1038/ncomms13219	Candidatus Jacksonbacteria
10.1038/ncomms13219	Candidatus Veblenbacteria
10.1038/ncomms13219	Candidatus Neelsonbacteria
10.1038/ncomms13219	Candidatus Colwellbacteria
10.1038/ncomms13219	Candidatus Liptonbacteria
10.1038/ncomms13219	Candidatus Harrisonbacteria
10.1038/ncomms13219	Candidatus Yonathbacteria
10.1038/ncomms13219	Candidatus Lloydbacteria
10.1038/ncomms13219	Candidatus Abawacabacteria
10.1038/ncomms13219	Candidatus Doudnabacteria
10.1038/ismej.2013.111	Candidatus Poribacteria
10.1111/1462-2920.13362	Candidatus Desantisbacteria
10.1038/nature12352	Candidatus Omnitrophica
10.1038/nature12352	Candidatus Aminicenantes
10.1126/science.1132690	Candidatus Micrarchaeota
10.1038/nature14486	Candidatus Magasanikbacteria
10.1073/pnas.0801980105	Candidatus Korarchaeota
10.1038/nature12352	Candidatus Fervidibacteria
10.1038/nature12352	Candidatus Aenigmarchaeota
10.1038/ismej.2016.43	Candidatus Fermentibacteria
10.1038/ismej.2013.174	Candidatus Bathyarchaeota
10.1016/j.cub.2015.01.014	Candidatus Woesearchaeota
10.1016/j.cub.2015.01.014	Candidatus Kryptonia
10.1038/nature12352	Candidatus Diapherotrites
10.1038/nature12352	Candidatus Latescibacteria
10.1038/nature21031	Candidatus Thorarchaeota
10.1038/ismej.2015.233	Candidatus Lindowbacteria
10.1038/ncomms13219	Candidatus Parvarchaeota
10.1038/nature12352	Candidatus Cloacimonetes
10.1038/nature12352	Candidatus Hydrogenedentes
10.1038/nature12352	Candidatus Acetothermia

10.1038/nature12352	Candidatus Nanohaloarchaeota
10.1038/ncomms13219	Candidatus Eisenbacteria
10.1186/s40168-015-0077-6	candidate division WOR-3
10.1038/nature21031	Lokiarchaeota
10.1038/nature21031	Odinarchaeota
10.1038/nature21031	Heimdallarchaeota

---

## 2.2 28

### 2.2.1 Bibliographic search for metagenomes

The reserarch for coral metagenomes started last year. The actual list is:

Tabela 2.2: table 1

IDs
mgm4440319.3
mgm4440370.3
mgm4440371.3
mgm4440372.3
mgm4440373.3
mgm4440374.3
mgm4440375.3
mgm4440376.3
mgm4440377.3
mgm4440378.3
mgm4440379.3
mgm4440380.3
mgm4440381.3
mgm4445755.3
mgm4445756.3
mgm4480739.3
mgm4480740.3
mgm4480741.3
mgm4480748.3
mgm4480750.3
mgm4487909.3
mgm4487910.3
mgm4487911.3
mgm4516541.3
mgm4516694.3
mgm4653307.3
mgm4694757.3
mgm4694758.3

mgm4694759.3  
mgm4694760.3  
SRR1275409  
SRR1275449  
SRR1283349  
SRR1283371  
SRR1283377  
SRR1283433  
SRR1283435  
SRR1283437  
SRR1286223  
SRR1286225  
SRR1286226  
SRR1286227  
SRR1286229  
SRR1286232  
SRR1822488  
SRR1822516  
SRR3499156  
SRR3569370  
SRR3694369  
SRR3694370  
SRR3694371  
SRR3694372  
SRR5215424  
SRR5215454  
SRR5215455  
SRR5215456  
SRR5215457  
SRR5215458  
SRR5215462  
SRR5605611

I found these metagenomes in the article: "Metagenomic analysis reveals a green sulfur bacterium as a potential coral symbiont"

SRR2937345  
SRR2937346  
SRR2937347  
SRR2937348  
SRR2937349  
SRR2937350  
SRR2937351  
SRR2937352

SRR2937353  
SRR2937354  
SRR2937355  
SRR2937356

Especie: *Platygyra carnosa* Healthy

I found other metagenomes of coral from article doi 10.3389/fmars.2018.00101 updated the file pmc\_results\_1.txt in the repository Lab\_book. I continue to look the articles in results. Estou atualizando a lista pmc\_results\_2.txt Na pesquisa bibliografica olhando o titulo ja me faz perceber se devo descartar e olhar. E olho aqueles que marquei para olhar. Ao olhar, leio o resumo procurando por metodos. E vou para os metodos do artigo para checar. Checking the sizes of metagenomes files. The mg-rast metagenomes base have 72 Gb.

The pipeline of bioinformatic is different for MGRAST and NCBI. The size of NCBI should be superestimated, because the ncbi says the file size of sra file, but most of them is paired-end metagenomes, so when we apply fastq-dump, its generate two files fastq.

# Capítulo 3

## Download of metagenomes

### 3.1 Download of mg-rast files

Espao no SDU Disponvel para o ebiodiv: 10Tb Bia: 5Tb Rilquer: 2T Remanescente: 3Tb

- Working folder: *scratch/ebiodiv/leticia.cavalcante/mg\_rast*

I insert the list of metagenomes in the files before using it. After this, I used the following command line:

- Command: *nohup bash download\_curl\_mgrast\_corais.sh & download\_curl\_mgrast\_corais.nohupout &*

### 3.2 Download of NCBI metagenomes

I use the script *download\_sra\_wget\_corais.sh*, *libs* folder. I used the *wget*, because the *curl* is getting some problem in SDU. I noted that the size of the files is different:

Tabela 3.1: Comparing sizes of files

ID of metagenome	the size in NCBI site	size of file in SDU
SRR6785058	317.00 Mb	318M
SRR6785057	364.00 Mb	365M
SRR6785056	560.00 Mb	561M
SRR6785055	624.00 Mb	625M

So I checked the others files:

A Bia me informou que o SDU arredonda os valores de tamanho dos arquivos, entao at o momento nao tive problemas com o download dos arquivos do *mg\_rast*

Tabela 3.2: Comparing sizes of files 2

ID of metagenome	the size in NCBI site	size of file in SDU	size of cleaned file
mgm4440319.3.299.1	30M	29.1 MB	28M
mgm4440370.3.299.1	3,6M	3.5 MB	3,5M
mgm4440371.3.299.1	5,0M	4.9 MB	4,8M
mgm4440372.3.299.1	6,0M	6.0 MB	5,9M
mgm4440373.3.299.1	6,2M	6.1 MB	6,0M
mgm4440374.3.299.1	4,1M	4.1 MB	4,0M
mgm4440375.3.299.1	3,8M	3.7 MB	3,7M
mgm4440376.3.299.1	3,9M	3.9 MB	3,8M
mgm4440377.3.299.1	3,5M	3.5 MB	3,4M
mgm4440378.3.299.1	6,2M	6.2 MB	6,1M
mgm4440379.3.299.1	7,0M	7.0 MB	6,9M
mgm4440380.3.299.1	5,2M	5.2 MB	5,2M
mgm4440381.3.299.1	6,4M	6.4 MB	6,4M
mgm4445755.3.299.1	158M	157.0 MB	155M
mgm4445756.3.299.1	150M	149.9 MB	147M
mgm4480739.3.299.1	8,0M	7.9 MB	7,9M
mgm4480740.3.299.1	12M	11.3 MB	12M
mgm4480741.3.299.1	8,5M	8.5 MB	8,5M
mgm4480742.3.299.1	10M	12.9 MB	10M
mgm4480743.3.299.1	15M	10.0 MB	14M
mgm4484839.3.299.1	13M	14.1 MB	13M
mgm4487909.3.299.1	17M	16.5 MB	17M
mgm4487910.3.299.1	36M	35.6 MB	36M
mgm4487911.3.299.1	12M	11.4 MB	12M
mgm4516541.3.299.1	161M	160.2 MB	163M
mgm4516694.3.299.1	193M	192.9 MB	193M
mgm4653307.3.299.1	17M	16.0 MB	17M
mgm4694757.3.299.1	1,9G	1.8 GB	1,9G
mgm4694758.3.299.1	2,2G	2.1 GB	2,2G
mgm4694759.3.299.1	1,7G	1.7 GB	1,8G
mgm4694760.3.299.1	592M	1.6 GB	597M

## Capítulo 4

# Format Conversion of NCBI metagenomes

Adaptei o script da Bia para fazer a conversao do dos arquivos .sra  
Inicialmente submeti apenas um na cpu\_dev para testar:

Script: *teste\_slurm\_job\_fastq\_dump\_corais.sh*  
Numero do job: *220896*

# Capítulo 5

## Quality filter

This step is only required for NCBI metagenomes. The command line was proposed by Bia:

- `trim_qual_left 25`
- `trim_qual_right 25`



# Capítulo 6

## Uniformity filter (size and N bases)

### 6.1 Command line

Parameters:

- `min_len` 80
- `ns_max_p` 2
- `out_format` 1

- Command: `nohup bash slurm_job_prinseq_single_corais_FASTA.bash &&`  
`slurm_prinseq_corais.out &`

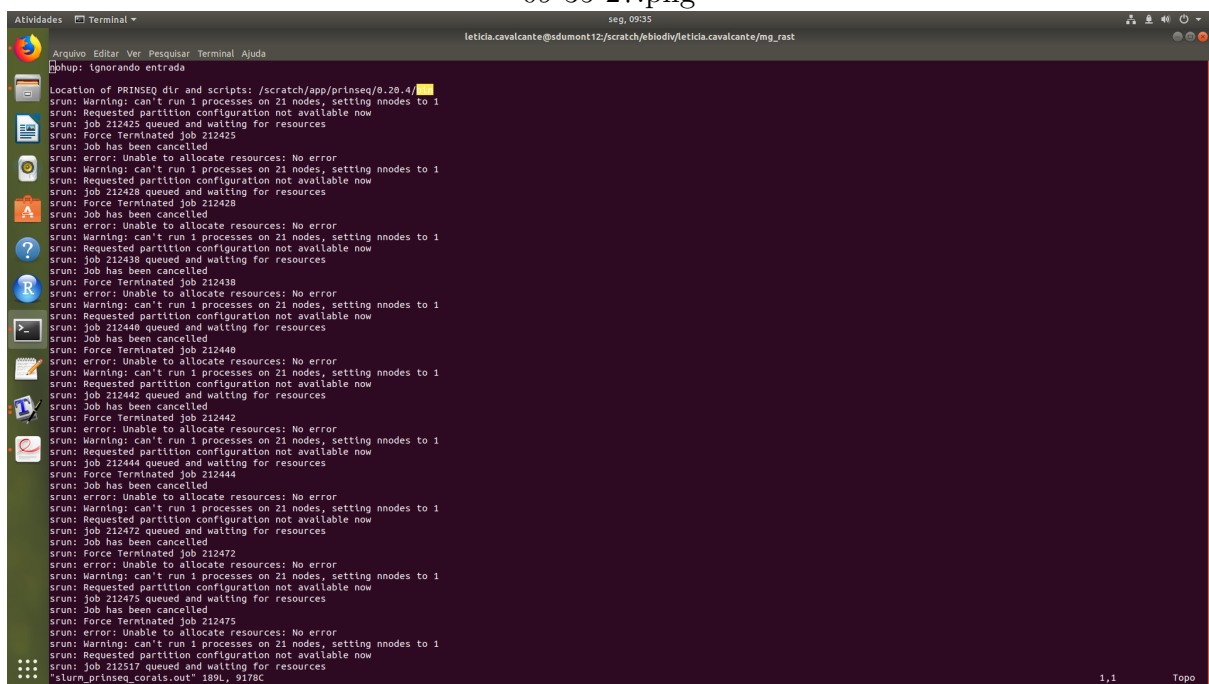
Deu erro o job nohup: ignorando entrada

Location of PRINSEQ dir and scripts: /scratch/app/prinseq/0.20.4/bin srun Warning: can't run 1 processes on 21 nodes, setting nnodes to 1 srun Requested partition configuration not available now srun job 212425 queued and waiting for resources srun Force Terminated job 212425 srun Job has been cancelled srun error: Unable to allocate resources: No error srun Warning: can't run 1 processes on 21 nodes, setting nnodes to 1 srun Requested partition configuration not available now srun job 212428 queued and waiting for resources srun Force Terminated job 212428 srun Job has been cancelled

Ressubmeti o job com:

- Command: `sbatch slurm_job_prinseq_single_corais_FASTA.bash`

09-35-27.png



```
Atividades Terminal seg, 09/35
Arquivo Editar Ver Pesquisar Terminal Ajuda
Jobup: Ignorando entrada
Location of PRINSEQ dir and scripts: /scratch/app/prinseq/0.20.4/
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212425 queued and waiting for resources
slurm: Force Terminated job 212425
slurm: Job has been cancelled
slurm: error: Unable to allocate resources: No error
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212428 queued and waiting for resources
slurm: Force Terminated job 212428
slurm: Job has been cancelled
slurm: error: Unable to allocate resources: No error
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212438 queued and waiting for resources
slurm: Force Terminated job 212438
slurm: error: Unable to allocate resources: No error
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212440 queued and waiting for resources
slurm: Job has been cancelled
slurm: Force Terminated job 212440
slurm: error: Unable to allocate resources: No error
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212442 queued and waiting for resources
slurm: Force Terminated job 212442
slurm: error: Unable to allocate resources: No error
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212444 queued and waiting for resources
slurm: Force Terminated job 212444
slurm: Job has been cancelled
slurm: error: Unable to allocate resources: No error
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212472 queued and waiting for resources
slurm: Force Terminated job 212472
slurm: error: Unable to allocate resources: No error
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212475 queued and waiting for resources
slurm: Force Terminated job 212475
slurm: error: Unable to allocate resources: No error
slurm: Warning: can't run 1 processes on 21 nodes, setting nnodes to 1
slurm: Requested partition configuration not available now
slurm: Job 212517 queued and waiting for resources
***
*** "slurm_prinseq_corals.out" 189L, 9178C
1,1 Topo
```

Figura 6.1: Erro no job no SDU

# Capítulo 7

## Profiling metagenomes

### 7.1 Mg-Rast metagenomes

I used the following script in the following folder:

```
- Folder: scratch/ebiodiv/leticia.cavalcante/mg_rast/filtered_prinseq_good  
- Command: sbatch slurm_job_kraken2_corais.sh
```

The job doesn't work, o erro aparece na proxima figura

Ressubmeti o job, modificando a localizacao da DB do Kraken para a home do Rilquer.

Numero do job: 216410

Esse problema foi resolvido modificando o endereco da base para o scratch do Rilquer.

### 7.2 Kraken-biom

Pasta onde est instalado kraken-biom:

`/home/leticia/.local/bin`

Para executar: `python2.7 .kraken-biom`

Executar o help do kraken-biom:

`kraken-biom -h`

Abrir no vim o arquivo `.bashrc` e inserir:

`export PATH=$PATH:/home/leticia/.local/bin/kraken-biom`

Executar o help do kraken-biom:

`kraken-biom -h`

Eu fiz um teste da etapa "Creation of BIOM table of abundances" da pipeline da bia com os seguintes passos: Na pasta `/home/leticia/Documentos/libs/leticia.profiling_metagenomes`:

- `kraken-biom selected_file -o table.biom --max D --min P`
- `biom convert -i table.biom -o table.from_biom_with_taxonomy.txt --to-tsv --header-key taxonomy`

```

Arquivo Editor Ver Pesquisar Terminal Ajuda
sex, 16:35
leticia.cavalcante@sdumont12/scratch/ebiodiv/leticia.cavalcante/mg_rast/filtered_prinseq_good

-rw-r--r-- 1 leticia.cavalcante ebiodiv 1,8G Set 12 20:41 mgm4694759_prinseq_good_4RbF.fasta
-rw-r--r-- 1 leticia.cavalcante ebiodiv 597M Set 12 20:42 mgm4694760_prinseq_good_B0zy.fasta
-rw-r--r-- 1 leticia.cavalcante ebiodiv 556 Set 13 14:53 slurm-215145.out
-rw-r--r-- 1 leticia.cavalcante ebiodiv 9,6K Set 13 16:54 slurm-215343.out
-rw-r--r-- 1 leticia.cavalcante ebiodiv 3,4K Set 13 15:21 slurm_job_kraken2_corais.sh
[leticia.cavalcante@sdumont12 filtered_prinseq_good]$ cat slurm-215343.out
sdumont1027-1029,1066,1094,1206-1207,1276-1277,1312-1318,1338-1339,1493-1494,5014-5023]
sdumont1027 sdumont1028 sdumont1029 sdumont1066 sdumont1094 sdumont1206 sdumont1207 sdumont1276 sdumont1277 sdumont1312 sdumont1313 sdumont1314 sdumont1315 sdumont1316 sdumont1317 sdu
mont1318 sdumont1338 sdumont1339 sdumont1493 sdumont1494 sdumont5014 sdumont5015 sdumont5016 sdumont5017 sdumont5018 sdumont5019 sdumont5020 sdumont5021 sdumont5022 sdumont5023
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440319_prinseq_good_DVfg.fasta!
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440370_prinseq_good_SiDP.fasta!
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440371_prinseq_good_EDLk.fasta!
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440372_prinseq_good_84wL.fasta!
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440373_prinseq_good_JqUa.fasta!
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440374_prinseq_good_9E3C.fasta!
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440375_prinseq_good_Lvcn.fasta!
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440376_prinseq_good_blxz.fasta!
srn: Warning: can't run 1 processes on 30 nodes, setting nnodes to 1
kraken2: database ("/prj/ebiodiv/maria.costa/Kraken2_DB") does not contain necessary file taxo.k2d
srn: error: sdumont1027: task 0: Exited with exit code 2
Produced profile and report for file: mgm4440377_prinseq_good_02aK.fasta!

```

Figura 7.1: 2o erro no job no SDU

- `perl filterRank.pl input table.from_biom_with_taxonomy.txt --rank p & abundance.matrix`

## 7.3 Teste com o kraken no scratch

Linha de teste:

`perl selectGroups.pl input mgm4440370_prinseq_good_SiDP.fasta_kraken.report --file_groups groups.txt & selected_file`

- First Command: `sbatch slurm_job_kraken2_corais.sh`
- Second Command:  
`kraken2 -db /prj/ebiodiv/rilquer.silva/Serrapilheira`  
`/Kraken2_custom_DB/ mgm4440370_prinseq_good_SiDP.fasta`  
`-output mgm4440370_prinseq_good_SiDP.fasta_kraken.profiled`  
`-use-names -report mgm4440370_prinseq_good_SiDP.fasta_kraken.report`

Ja testei o comando acima na home do SDU e agora no scratch

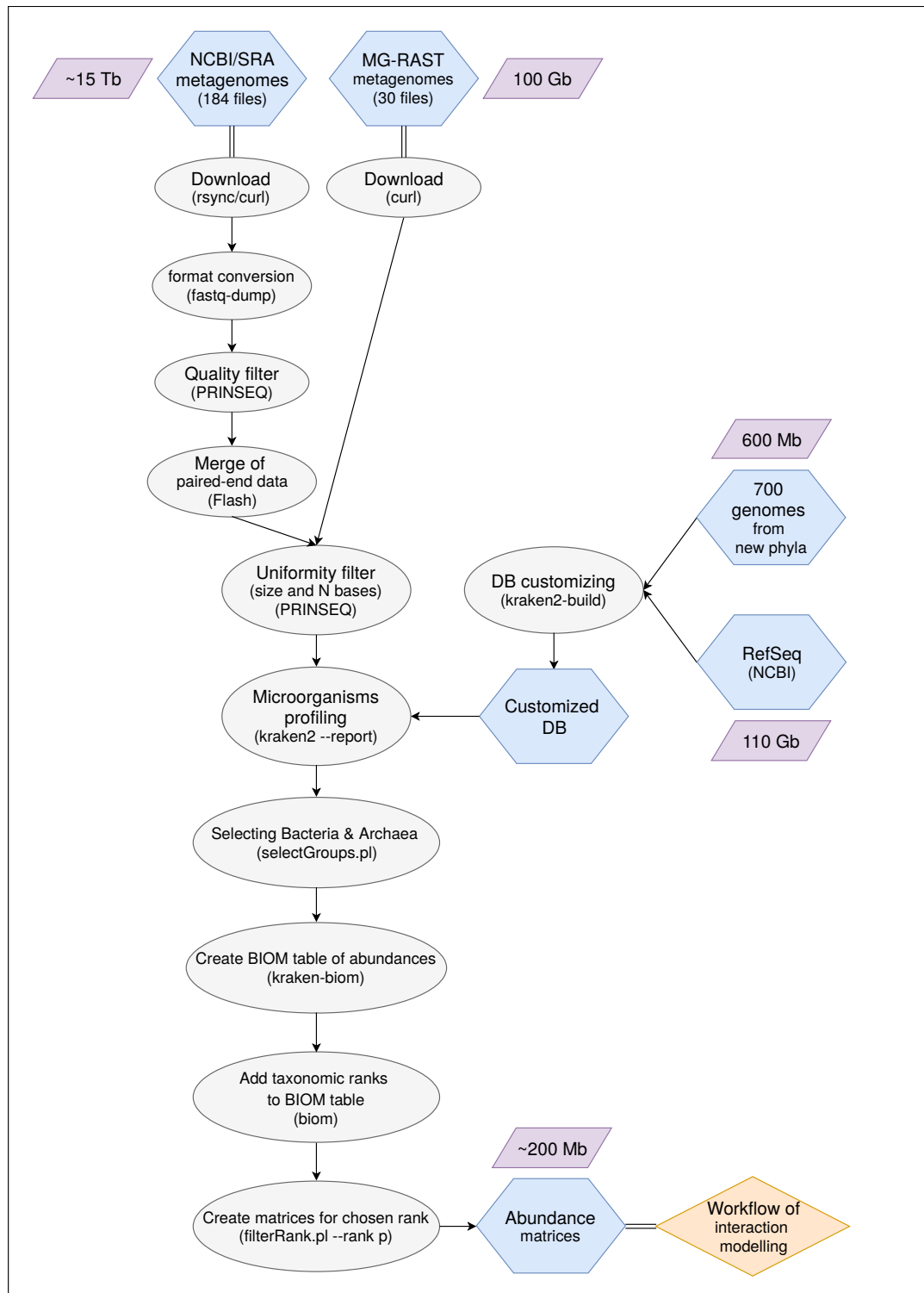


Figura 7.2: Pipeline of taxonomic annotation

## Capítulo 8

# Functional annotation of metagenomes

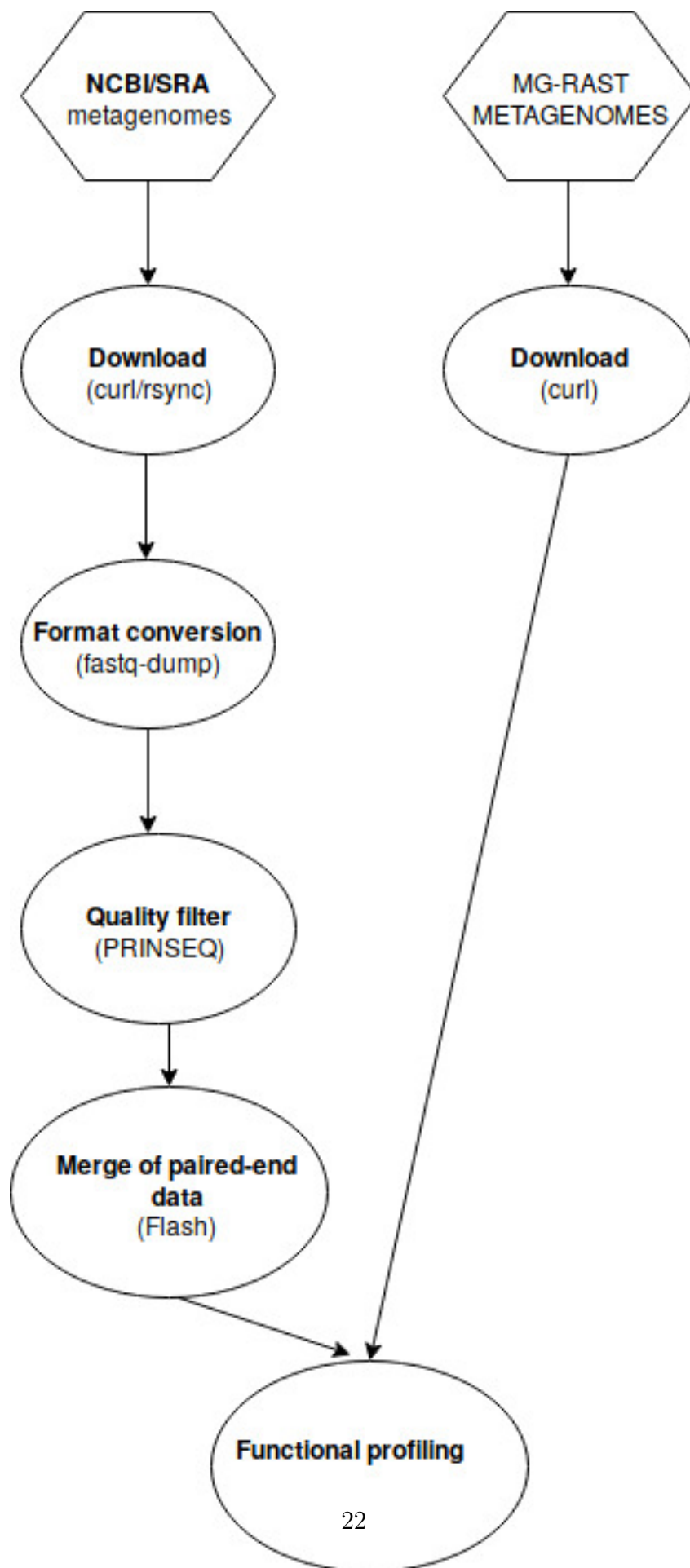


Figura 8.1: Pipeline of functional annotation

# Capítulo 9

## references

Articles list:

- 10.1371/journal.pone.0071301: Relata resultados que eu acreditava ter sido a primeira a encontrar
- 10.1038/nature14486: reconstruction of microorganism's genomes we use
- 10.1038/nmicrobiol.2016.48: three of life, including the Candidate Phyla Radiation
- 10.1146/annurev.micro.57.030502.090759: speaks about the uncultured majority of microorganisms
- 10.1038/ismej.2016.174: revision of rare biosphere
- 10.1038/nrmicro3400: another revision of rare biosphere
- 10.1126/science.1224041: metabolic activities of Candidatus Parcubacteria, one of super-phyla of CPR
- 10.1128/MMBR.00009-08: Revision of bioinformatic methods and steps for meta-genomic
- 10.1186/s40168-018-0428-1: Sponge as holobiont. Note: This article has a important information about microbial ecology: "Network and modeling analyses aim to disentangle the strength and nature (positive, negative, or neutral) of the interactions and predict their dynamics. Bacteria-bacteria network analysis of the core microbiota in different sponge species has revealed a low connective network with very few strong and many weak unidirectional interactions (i.e., amensalism [/0] and commensalism [+ /0] prevailed over cooperation [+ /+] and competition [/]. These findings are consistent with mathematical models that predict that weak and non-cooperative interactions help to stabilize highly diverse microbial communities, whereas cooperation yields instability in the long term by fueling positive feedbacks"
- 10.1016/j.tim.2009.09.004: Microbial disease and the coral holobiont
- 10.3389/fmicb.2017.00618: Comparative Metagenomics of the Polymicrobial Black Band Disease of Corals



- 10.1038/nrmicro1643: The role of ecological theory in microbial ecology
- 10.1038/nrmicro3218: Explaining microbial genomic diversity in light of evolutionary ecology
- 10.1111/j.1462-2920.2009.01935.x: Metagenomic analysis of stressed coral holobionts
- 10.1038/nature06810: Functional metagenomic profiling of nine biomes
- 10.3389/fcimb.2014.00176: Microbes in the coral holobiont: partners through evolution, development, and ecological interactions
- 10.1038/ismej.2015.39: The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts
- 10.1111/j.1462-2920.2007.01383.x: Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*
- 10.1038/nrmicrobiol.2015.32: Metagenomics uncovers gaps in amplicon-based detection of microbial diversity
- 10.1038/ismej.2016.45: Challenges in microbial ecology: building predictive understanding of community function and dynamics
- 10.1111/j.1462-2920.2009.02113.x: Microbial functional structure of *Montastraea faveolata*, an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies
- 10.1111/j.1758-2229.2010.00234.x:

# Capítulo 10

## Softwares, instalacao e linhas

### 10.1 Profilling metagenomes

Instalar o kraken-biome

- Folder: */home/leticia*
- Command: *pip install kraken-biom*
- Site: *<https://github.com/smdabdoub/kraken-biom>*

Para atualizacao:  
git commit  
git push origin master

Para transferencia:  
maquina remota para local: `scp leticia.cavalcante@login.sdumont.lncc.br:/scratch/ebiodiv/leticia.cavalcante:/home/leticia/Documentos/dados`

## Referências Bibliográficas