

#### Advanced Linux Usage

2021-05-18

Martin Dahlö martin.dahlo@uppmax.uu.se

Enabler for Life Sciences











#### Same program, many files

```
$ ls -l
total 0
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_1.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_2.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_3.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_4.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_5.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_6.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_7.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_8.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep   1 16:42 sample_9.bam
$ my_prog sample_1.bam
```



#### Same program, many files



#### Same program, many files

```
$ 1s -1
total 0
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 1.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 2.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 3.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 4.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 5.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 6.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 7.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 8.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 16:42 sample 9.bam
$ my proq sample 1.bam
$ my prog sample 2.bam
$ my prog sample 3.bam
$ my prog sample 4.bam
$ my prog sample 5.bam
$ my prog sample 6.bam
$ my prog sample 7.bam
$ my prog sample 8.bam
$ my prog sample 9.bam
```



- Same program, many files
  - 10 files? Ok
  - 1000 files? Not ok



- Same program, many files
  - 10 files? Ok
  - 1000 files? Not ok
- Reproducibility
  - Self and others



- Same program, many files
  - 10 files? Ok
  - 1000 files? Not ok
- Reproducibility
  - Self and others

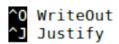
A solution - write a script!

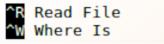


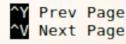
```
total 0
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_1.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_2.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_3.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_4.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_5.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_6.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_7.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_8.bam
-rw-rw-r-- 1 dahlo dahlo 0 Sep 1 17:18 sample_9.bam
shano analysis.sh
```

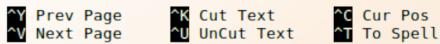


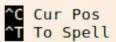
GNU nano 2.0.9 File: analysis.sh











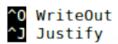


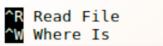
GNU nano 2.0.9

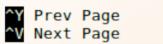
File: analysis.sh

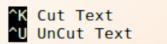
Modified

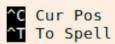
my prog sample 1.bam













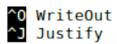
GNU nano 2.0.9

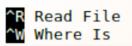
File: analysis.sh

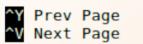
Modified

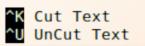
my\_prog sample\_1.bam
my\_prog sample 2.bam

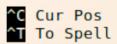
^G Get Help ^X Exit







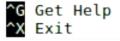


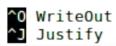


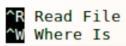


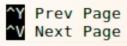
GNU nano 2.0.9 File: analysis.sh Modified

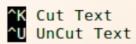
my\_prog sample\_1.bam
my\_prog sample\_2.bam
my\_prog sample\_3.bam
my\_prog sample\_4.bam
my\_prog sample\_5.bam
my\_prog sample\_6.bam
my\_prog sample\_7.bam
my\_prog sample\_8.bam
my\_prog sample\_9.bam

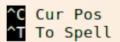














```
$ l
total 4.0K
-rw-rw-r-- 1 dahlo dahlo 267 Sep 7 09:34 analysis.sh
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 1.bam
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 2.bam
                          0 Sep 1 17:18 sample 3.bam
-rw-rw-r-- 1 dahlo dahlo
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 4.bam
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 5.bam
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 6.bam
                           0 Sep 1 17:18 sample 7.bam
-rw-rw-r-- 1 dahlo dahlo
                           0 Sep 1 17:18 sample 8.bam
-rw-rw-r-- 1 dahlo dahlo
-rw-rw-r-- 1 dahlo dahlo
                           0 Sep 1 17:18 sample 9.bam
Ś
```



```
$ l
total 4.0K
-rw-rw-r-- 1 dahlo dahlo 267 Sep 7 09:34 analysis.sh
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 1.bam
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 2.bam
                          0 Sep 1 17:18 sample 3.bam
-rw-rw-r-- 1 dahlo dahlo
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 4.bam
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 5.bam
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 6.bam
                           0 Sep 1 17:18 sample 7.bam
-rw-rw-r-- 1 dahlo dahlo
                           0 Sep 1 17:18 sample 8.bam
-rw-rw-r-- 1 dahlo dahlo
-rw-rw-r-- 1 dahlo dahlo
                           0 Sep 1 17:18 sample 9.bam
$ bash analysis.sh
```



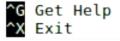
```
$ 1
total 4.0K
-rw-rw-r-- 1 dahlo dahlo 267 Sep 7 09:34 analysis.sh
                          0 Sep 1 17:18 sample 1.bam
-rw-rw-r-- 1 dahlo dahlo
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 2.bam
                          0 Sep 1 17:18 sample 3.bam
-rw-rw-r-- 1 dahlo dahlo
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 4.bam
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 5.bam
-rw-rw-r-- 1 dahlo dahlo
                          0 Sep 1 17:18 sample 6.bam
                           0 Sep 1 17:18 sample 7.bam
-rw-rw-r-- 1 dahlo dahlo
                           0 Sep 1 17:18 sample 8.bam
-rw-rw-r-- 1 dahlo dahlo
-rw-rw-r-- 1 dahlo dahlo
                           0 Sep 1 17:18 sample 9.bam
$ bash analysis.sh
```

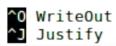
Still not OK for 1000 or more files!

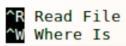


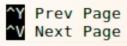
GNU nano 2.0.9 File: analysis.sh Modified

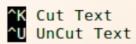
my\_prog sample\_1.bam
my\_prog sample\_2.bam
my\_prog sample\_3.bam
my\_prog sample\_4.bam
my\_prog sample\_5.bam
my\_prog sample\_6.bam
my\_prog sample\_7.bam
my\_prog sample\_8.bam
my\_prog sample\_9.bam

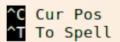










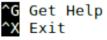


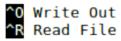


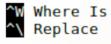
GNU nano 2.5.3

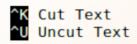
```
File: analysis.sh
```

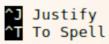
```
my_prog -r references/human_genome.fa sample_1.bam
my_prog -r references/human_genome.fa sample_2.bam
my_prog -r references/human_genome.fa sample_3.bam
my_prog -r references/human_genome.fa sample_4.bam
my_prog -r references/human_genome.fa sample_5.bam
my_prog -r references/human_genome.fa sample_6.bam
my_prog -r references/human_genome.fa sample_7.bam
my_prog -r references/human_genome.fa sample_8.bam
my_prog -r references/human_genome.fa sample_9.bam
```

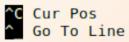


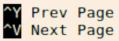














### Assigning

```
my_variable=5
my_variable="nice text"
```



#### Assigning

```
my_variable=5
my_variable="nice text"
```

#### Using



#### Assigning

```
my_variable=5
my_variable="nice text"
```

#### Using

```
$ my_variable="Dave"
```



#### Assigning

```
my_variable=5
my_variable="nice text"
```

#### Using

```
$ my_variable="Dave"
$ echo "Hello, $my_variable."
```



#### Assigning

```
my_variable=5
my_variable="nice text"
```

#### Using

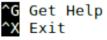
```
$ my_variable="Dave"
$ echo "Hello, $my_variable."
Hello, Dave.
```

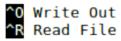


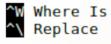
GNU nano 2.5.3

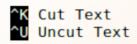
```
File: analysis.sh
```

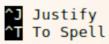
```
my_prog -r references/human_genome.fa sample_1.bam
my_prog -r references/human_genome.fa sample_2.bam
my_prog -r references/human_genome.fa sample_3.bam
my_prog -r references/human_genome.fa sample_4.bam
my_prog -r references/human_genome.fa sample_5.bam
my_prog -r references/human_genome.fa sample_6.bam
my_prog -r references/human_genome.fa sample_7.bam
my_prog -r references/human_genome.fa sample_8.bam
my_prog -r references/human_genome.fa sample_9.bam
```

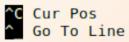


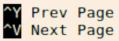












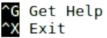


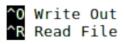
GNU nano 2.5.3

File: analysis.sh

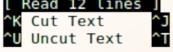
```
ref=references/human_genome.fa
```

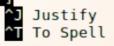
```
my_prog -r $ref sample_1.bam
my_prog -r $ref sample_2.bam
my_prog -r $ref sample_3.bam
my_prog -r $ref sample_4.bam
my_prog -r $ref sample_5.bam
my_prog -r $ref sample_6.bam
my_prog -r $ref sample_7.bam
my_prog -r $ref sample_8.bam
my_prog -r $ref sample_8.bam
my_prog -r $ref sample_9.bam
```



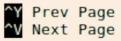












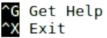


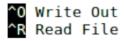
GNU nano 2.5.3

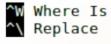
File: analysis.sh

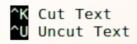
```
ref=references/goat_genome_version4.1.fa
```

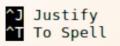
```
my_prog -r $ref sample_1.bam
my_prog -r $ref sample_2.bam
my_prog -r $ref sample_3.bam
my_prog -r $ref sample_4.bam
my_prog -r $ref sample_5.bam
my_prog -r $ref sample_6.bam
my_prog -r $ref sample_7.bam
my_prog -r $ref sample_8.bam
my_prog -r $ref sample_8.bam
my_prog -r $ref sample_9.bam
```



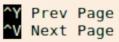














```
for var in 1 2 3;
do
    echo $var
done
```

```
$ bash loop_test.sh
1
2
3
```



```
for var in text works too;
do
    echo $var
done
```

```
$ bash loop_test.sh
text
works
too
$
```



```
for var in mix them 5;
do
    echo $var
done
```

```
$ bash loop_test.sh
mix
them
5
```



```
for var in *.txt;
do
    echo $var
done
```

```
$ bash loop_test.sh
all.txt
examples.txt
readme.txt
```

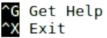


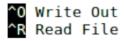
GNU nano 2.5.3

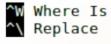
File: analysis.sh

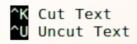
```
ref=references/goat_genome_version4.1.fa
```

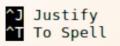
```
my_prog -r $ref sample_1.bam
my_prog -r $ref sample_2.bam
my_prog -r $ref sample_3.bam
my_prog -r $ref sample_4.bam
my_prog -r $ref sample_5.bam
my_prog -r $ref sample_6.bam
my_prog -r $ref sample_7.bam
my_prog -r $ref sample_8.bam
my_prog -r $ref sample_8.bam
my_prog -r $ref sample_9.bam
```



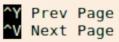










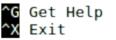


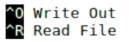


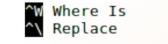
GNU nano 2.5.3

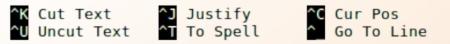
File: analysis.sh

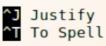
```
ref=references/goat genome version4.1.fa
for file in *.bam ;
   my prog -r $ref $file
done
```



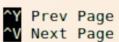












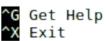


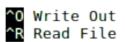
GNU nano 2.5.3

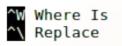
File: analysis.sh

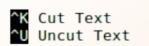
```
ref=references/goat genome version4.1.fa
for file in *.bam ;
   echo my prog -r $ref $file
done
```

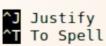
#### Debugging!



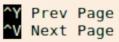














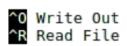
GNU nano 2.5.3

File: analysis.sh

```
ref=references/goat_genome_version4.1.fa
for file in *.bam ;
do
    echo my_prog -r $ref $file
done
```

```
$ bash analysis.sh
my_prog -r references/goat_genome_version4.1.fa sample_1.bam
my_prog -r references/goat_genome_version4.1.fa sample_2.bam
my_prog -r references/goat_genome_version4.1.fa sample_3.bam
my_prog -r references/goat_genome_version4.1.fa sample_4.bam
my_prog -r references/goat_genome_version4.1.fa sample_5.bam
my_prog -r references/goat_genome_version4.1.fa sample_6.bam
my_prog -r references/goat_genome_version4.1.fa sample_7.bam
my_prog -r references/goat_genome_version4.1.fa sample_8.bam
my_prog -r references/goat_genome_version4.1.fa sample_9.bam
```

```
^G Get Help
^X Exit
```





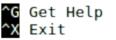


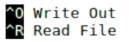


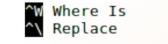
GNU nano 2.5.3

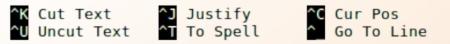
File: analysis.sh

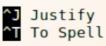
```
ref=references/goat genome version4.1.fa
for file in *.bam ;
   my prog -r $ref $file
done
```



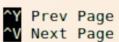














## **Arguments**

\$ bash analysis.sh



\$ bash analysis.sh data/



\$ bash analysis.sh data/

\$1



\$ bash analysis.sh data/ second\_argument

\$1 \$2



\$ bash analysis.sh data/ second\_argument third

\$1

\$2

\$3



\$ bash analysis.sh data/ second\_argument third "fourth argument"

\$1

\$2

\$3

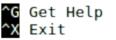
\$4

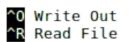


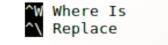
GNU nano 2.5.3

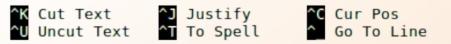
File: analysis.sh

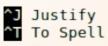
```
ref=references/goat genome version4.1.fa
for file in *.bam ;
   my_prog -r $ref $file
done
```



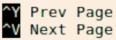










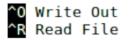


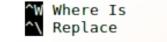


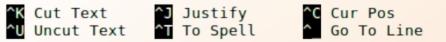
GNU nano 2.5.3

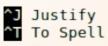
File: analysis.sh

```
ref=references/goat genome version4.1.fa
for file in $1/*.bam ;
   my prog -r $ref $file
done
```

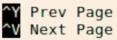














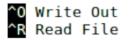
```
$ cat file.list
sample_1.bam
sample_3.bam
smaple_9.bam
```

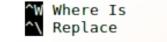


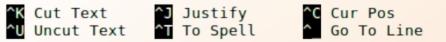
GNU nano 2.5.3

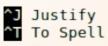
File: analysis.sh

```
ref=references/goat genome version4.1.fa
for file in $1/*.bam ;
   my prog -r $ref $file
done
```

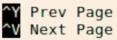










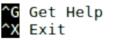


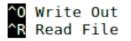


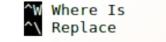
GNU nano 2.5.3

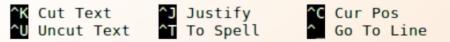
File: analysis.sh

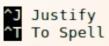
ref=references/goat genome version4.1.fa for file in \$( cat \$1 ); my\_prog -r \$ref \$file done



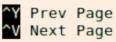














#### GNU nano 2.5.3

File: analysis.sh

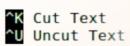
```
ref=references/goat_genome_version4.1.fa
for file in $( cat $1 ) ;
do
    my_prog -r $ref $file
done
```

```
$ cat file.list
sample_1.bam
sample_3.bam
sample_9.bam
$ bash analysis.sh file.list
```

```
^G Get Help
^X Exit
```

```
^O Write Out
^R Read File
```

```
^W Where Is
^\ Replace
```



```
if condition; then
  action
```

fi



```
if true; then
  echo "This is true"
fi
```

result:
This is true

```
if false; then
  echo "This is true"
fi
```

result:



```
if [[ 5 < 9 ]]; then
  echo "This is true"
fi</pre>
```

result:

This is true



```
if [[ 5 > 9 ]]; then
  echo "This is true"
fi
```

result:



```
if [[ 5 == 9 ]]; then
  echo "This is true"
fi
```

result:



```
if [["Hello" == "Hello"]]; then
  echo "This is true"
fi
```

result:

This is true

```
if [["Hello" == "Hi" ]]; then
  echo "This is true"
fi
```

result:



```
if [[ "Hello" == "Hel"* ]]; then
  echo "This is true"
fi
```

result:

This is true



```
for file in $1/*.bam ;
do
    echo my_prog $file
done
```



```
for file in $1/*.bam ;
do
    if [[ ... != "dog"* ]]; then
        echo my_prog $file
    fi
done
```

```
for file in $1/*.bam ;
do
    if [[ ... != "dog"* ]]; then
        echo my_prog $file
    fi
done
```

Ex: \$file is /path/to/dog\_1.bam

```
for file in $1/*.bam ;
do
    if [[ ... != "dog"* ]]; then
       echo my_prog $file
    fi
done
```

Ex: \$file is /path/to/dog\_1.bam

basename \$file

```
for file in $1/*.bam ;
do
    if [[ ... != "dog"* ]]; then
       echo my_prog $file
    fi
done
```

Ex: \$file is /path/to/dog\_1.bam

basename \$file

dog\_1.bam



```
for file in $1/*.bam ;
do
    if [[ $(basename $file) != "dog"* ]]; then
        echo my_prog $file
    fi
done
```

Ex: \$file is /path/to/dog\_1.bam

basename \$file

dog\_1.bam



```
for file in $1/*.bam ;
do
    if [[ $(basename $file) != "dog"* ]]; then
        my_prog $file
    fi
done
```

Ex: \$file is /path/to/dog\_1.bam

basename \$file

dog\_1.bam



- Programming is programming
  - Perl, Python, Bash, and more



- Programming is programming
  - Perl, Python, Bash, and more

```
for file in $1/*.bam ;
do
    if [[ $(basename $file) != "dog"* ]]; then
        my_prog $file
    fi
done
```



- Programming is programming
  - Perl, Python, Bash, and more



- Programming is programming
  - Perl, Python, Bash, and more



- Programming is programming
  - Perl, Python, Bash, and more
- Start with one, git gud, (learn another)



- Programming is programming
  - Perl, Python, Bash, and more
- Start with one, git gud, (learn another)

PYTHON



- Graphical text editor more similar to what you might be used to
- Launch through command line:

```
$ gedit
```



INS

```
test.py
~/work/tmp
  Open ▼
                                                                                            _ 0 ×
                                                                               Save
A = "ACGT"
if len(A) <= 3:
    print("Sequence A is smaller or equal than 3.")
elif (len(A) > 3) and (len(A) < 5):
    print("Sequence A is greater than 3 and smaller than 5.")
elif len(A) == 5:
    print("Sequence A is equal to 5.")
else:
    print("Sequence A is greater than 5.")
                                                         Python ▼ Tab Width: 4 ▼
                                                                                   Ln 4, Col 7
```



```
test.py
~/work/tmp
                                                                                       =
  Open ▼
                                                                                Save
1 A = "ACGT"
2 if len(A) <= 3:
      print("Sequence A is smaller or equal than 3.")
4 elif (len(A) > 3) and (len(A) < 5):
      print("Sequence A is greater than 3 and smaller than 5.")
6 elif len(A) == 5:
      print("Sequence A is equal to 5.")
8 else:
      print("Sequence A is greater than 5.")
9
10
11
                                                         Python ▼ Tab Width: 4 ▼
                                                                                    Ln 4, Col 7
                                                                                                    INS
```



```
test.py
~/work/tmp
                                                                                      =
  Open ▼
                                                                               Save
 1 A = "ACGT"
 2 if len(A) <= 3:
      print("Sequence A is smaller or equal than 3.")
 4 elif (len(A) > 3) and (len(A) < 5):
      print("Sequence A is greater than 3 and smaller than 5.")
 6 elif len(A) == 5:
      print("Sequence A is equal to 5.")
 8 else:
      print("Sequence A is greater than 5.")
11
```

Theme: Kate



```
test.py
~/work/tmp
Open ▼
                                                                             Save
A = "ACGT"
if len(A) <= 3:
     print("Sequence A is smaller or equal than 3.")
elif (len(A) > 3) and (len(A) < 5):
     print("Sequence A is greater than 3 and smaller than 5.")
elif len(A) == 5:
     print("Sequence A is equal to 5.")
else:
     print("Sequence A is greater than 5.")
                                                      Python ▼ Tab Width: 4 ▼
                                                                                 Ln 4, Col 7
                                                                                             ▼ INS
```

Theme: Oblivion



```
test.py
~/work/tmp
                                                                             Save
  Open ▼
 1 A = "ACGT"
2 if len(A) <= 3:
3 print("Sequence A is smaller or equal than 3.")
4 elif (len(A) > 3) and (len(A) < 5):
5 print("Sequence A is greater than 3 and smaller than 5.")
6 elif len(A) == 5:
    print("Sequence A is equal to 5.")
8 else:
      print("Sequence A is greater than 5.")
10
11
                                                       Python ▼ Tab Width: 4 ▼
                                                                                Ln 4, Col 7
                                                                                          ▼ INS
```

Theme: Monokai



- Menu Preferences View
  - Display line numbers
  - Display overview map
  - Highlight current line
  - Highlight matching brackets
- Menu Preferences Editor
  - Tab width 4
  - Insert spaces instead of tabs
- Menu Preferences Fonts & Colors
  - Kate or Oblivion



### Laboratory time once again!

https://nbisweden.github.io/workshop-ngsintro/2105/lab\_linux\_advanced.html