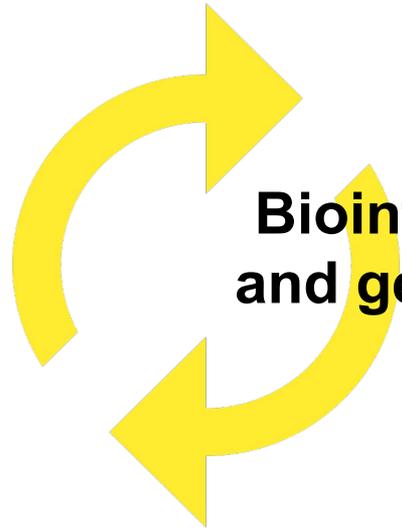


## Modules de formation 2019





**Bioinformatics platform dedicated to the genetics and genomics of tropical and Mediterranean plants and their pathogens**

genome assembly SNP detection  
phylogeny structural variation  
comparative genomics transcriptome assembly differential expression  
GWAS pangenomics  
population genetics metagenomics  
polyploidy



Rice



Banana



Palm



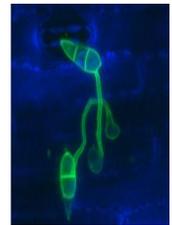
Sorghum



Coffee



Cassava



Magnaporthe



Larmande Pierre  
Sabot François  
Tando Ndomassi  
**Tranchant-Dubreuil  
Christine**



Comte Aurore  
Dereeper Alexis



**Orjuela-Bouniol Julie**



Bocs Stephanie  
De Lamotte Frédéric  
**Droc Gaetan**  
Dufayard Jean-François  
Hamelin Chantal  
Martin Guillaume  
Pitollat Bertrand  
**Ruiz Manuel**  
**Sarah Gautier**  
Summo Marilyne



**Rouard Mathieu**  
Guignon Valentin  
Catherine Breton



**Mahé Frédéric**  
**Ravel Sébastien**



Sempere Guilhem



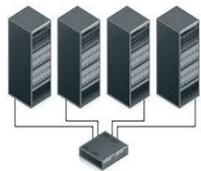
## Workflow manager

TOOLBOX  
Toolbox for generic NGS analyses

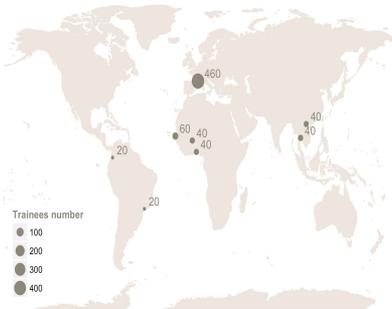
●●●●●  
SNAKEMAKE

Galaxy

## HPC and trainings....



37 courses organized last 7 years



IRD  
Institut de Recherche  
pour le Développement

cirad

## Genome Hubs & Information System



Gigwa

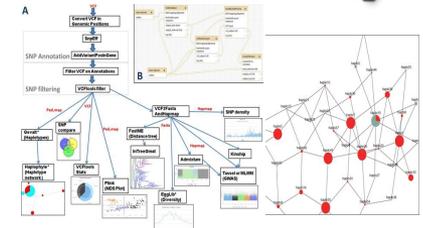
SNPs and Indels

GreenPhyl

Family Id	Family Name	Number of sequences	Status
EP000010	Cytoschrome P450 superfamily	6942	●●●
EP000017	AP2/ERF/ERF transcription factor family: ERF/ERF group (partial)	5142	●●●
EP000020	NAC transcription factor family	4574	●●●
EP000028	MADS transcription factor family		
EP000018	Hamam peroxidase superfamily		
EP000066	General substrate transporter superfamily		
EP000022	Subtilisin-like Serine Proteases family		
EP000019	NPF, NRT/PTIR FAMILY		

Gene families

SNIPlay



<https://github.com/SouthGreenPlatform>



@green\_bioinfo

- 18-19/03 — ● Guide de survie à Linux - IRD
- 21/03 — ● Initiation à l'utilisation du cluster CIRAD – CIRAD
- 22/03 — ● Initiation à l'utilisation du cluster itrop - IRD
- 15-16/04 — ● Initiation au gestionnaires de workflow SG & Gigwa – IRD
- 18-19/04 — ● Guide du Jedi en Linux & bash - CIRAD
- 13–16/05 — ● Python - IRD
- 17/05 — ● Initiation aux analyses de données transcriptomiques – IRD
- 21/05 — ● Utilisation avancée du cluster IRD – IRD
- 23-24/05 — ● Initiation aux analyses de données métagénomiques – IRD
- 6/06 — ● Manipulation de données et figures sous R – CIRAD
- 26-28/06 — ● Assemblage et annotation de transcriptomes - IRD

# Modules de formation 2019

- Toutes nos formations :  
<https://southgreenplatform.github.io/trainings/>
- Topo & TP : [Linux For Dummies](#)
- Environnement de travail : [Logiciels à installer](#)

# Guide de survie à Linux



[www.southgreen.fr](http://www.southgreen.fr)

<https://southgreenplatform.github.io/trainings>



## The objectif!

Lancez vos analyses bioinformatiques sous Linux !



## Applications

Connaître les principales commandes sous Linux

- Se déplacer dans l'arborescence de fichier : *pwd*, *ls*, *cd*, *mkdir* etc.
- Se connecter à un serveur, transférer les données : *ssh*, *scp*, *wget*
- Manipuler des fichiers : *head*, *tail*, *sort*, *cut*, *wc*, *grep*
- Lancer des logiciels bioinformatiques en ligne de commande



# Introduction

- **Systeme d'exploitation** réputé pour :
  - sa sécurité
  - ses mises à jour fréquentes
  - son prix et ses programmes gratuits

- Créé en 1991 par *Linus Torvalds*

- Basé sur l'OS propriétaire Unix (1969)

- Linux *gratuit* et *libre*

on peut avoir le code source, la "recette de fabrication"  
on peut copier, modifier, redistribuer



- **Système robuste et multi-plateforme**
- **Système multi-utilisateurs**

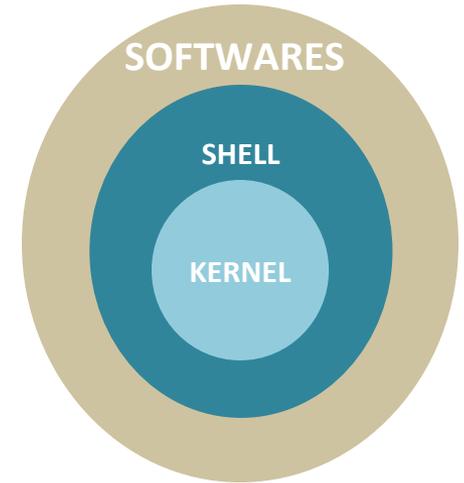
Plusieurs utilisateurs peuvent travailler en même temps

- **Système multi-tâches (processus/programmes)**

Chaque utilisateur peut lancer plusieurs programmes en même temps



**Distribution** : Noyau/Coeur + logiciels/programmes



- 2 façons d'utiliser linux :

en *mode graphique*



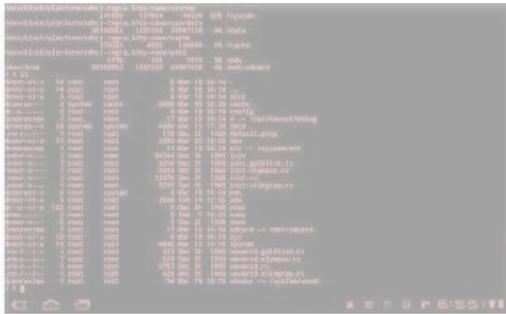




- Nombreux programmes rapides & puissants
- Facile de lier des commandes/programmes entre eux (workflow)
- Nombreux outils bioinformatique disponibles
- 90% des serveurs fonctionnent sous Linux



- Nombreux programmes rapides & puissants
- Facile de lier des commandes/programmes entre eux (workflow)
- Nombreux outils bioinformatique disponibles
- 90% des serveurs fonctionnent sous Linux



**Pas d'interfaces  
graphiques**

**Convivialité de la ligne  
de commande ?**





Nécessité de la pratique et de l'expérience

↔ Investissement non négligeable pour de bons résultats rapidement



# Environnement de travail

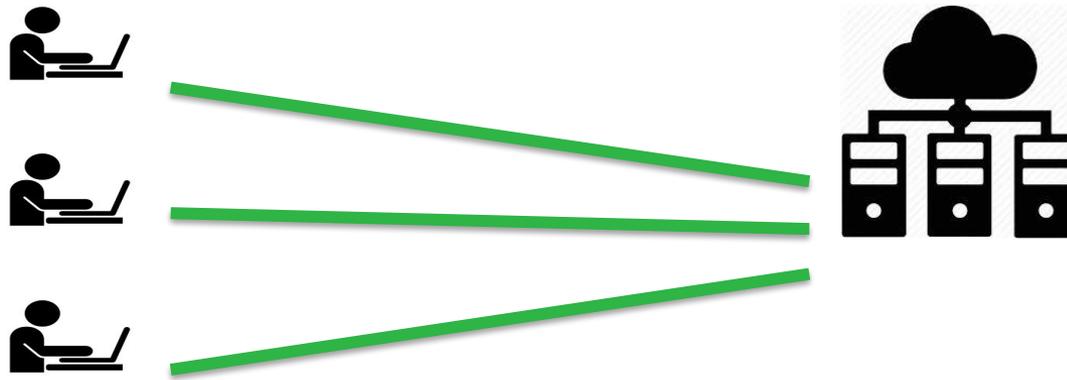
*Comment copier un fichier de son PC sur  
le serveur ?*



# Copier un fichier de son PC sur le serveur ?



- En se connectant sur un serveur linux distant de son ordinateur via le **protocole sftp**



## HPC South Green

- itrop (IRD) [bioinfo-nas.ird.fr](http://bioinfo-nas.ird.fr)
- HPC AGAP (CIRAD) [cc2.login.cirad.fr](http://cc2.login.cirad.fr)



# Filezilla

The screenshot shows the FileZilla application window. At the top, there is a toolbar and a header bar with the title "FileZilla". Below the header, there are input fields for "Hôte :", "Identifiant :", "Mot de passe :", and "Port :", along with a "Connexion rapide" dropdown menu.

The main window displays the "Gestionnaire de Sites" (Site Manager) dialog box. The dialog has a "Site local" field set to "/U:". The main area is titled "Sélectionnez une entrée :" and shows a tree view under "Mes Sites" with entries: "Nouveau site", "bioinfo-inter", and "nas". The "nas" entry is selected and highlighted with a blue box, labeled with a blue "2".

At the bottom of the dialog, there are several buttons: "Nouveau Site" (labeled with a blue "1"), "Nouveau Dossier", "Nouveau Favori", "Renommer", "Supprimer", and "Dupliquer".

On the right side of the dialog, there are tabs for "Général", "Avancé", and "Paramètres de trans". The "Général" tab is active, showing fields for "Hôte :", "Protocole :", "Type d'authentification :", "Identifiant :", and "Mot de passe :". The "Hôte" field contains "bioinfo-nas.ird.fr", the "Protocole" is "SFTP - SSH File Tran...", the "Type d'authentification" is "Normale", the "Identifiant" is "tranchant", and the "Mot de passe" is masked with dots. A blue box labeled "3" encompasses these fields.

At the bottom right of the dialog, there is a "Couleur de fond" dropdown set to "Aucune" and a "Commentaires" text area. A blue box labeled "4" encompasses the "Connexion" button at the bottom right of the dialog.



# Filezilla

**Quick Links**

The screenshot shows the Filezilla version 2.2.3 interface. At the top, the menu bar includes File, Edit, Transfer, View, Queue, Server, and Help. Below the menu bar is the Quick connect bar, which contains fields for Address (yourdomain.com), User (yourdomain), Password (masked with asterisks), and Port (21), along with a Quick connect dropdown button. The interface is divided into several panes: a Local Site pane on the left showing a tree view of local folders; a Remote Site pane on the right, currently empty; a Computer File List pane at the bottom left showing a detailed list of local files and folders; and a Queue Manager pane at the bottom, which is currently empty. The status bar at the bottom left shows 'Ready' and the bottom right shows 'Queue: 0 bytes'.

**Quick connect bar**

**Local Site**

**Remote Site**

**Computer File List**

**Queue Manager**



# Practice

filezilla, sftp

1

Go to [Practice 1](#) on our github



# Environnement de travail

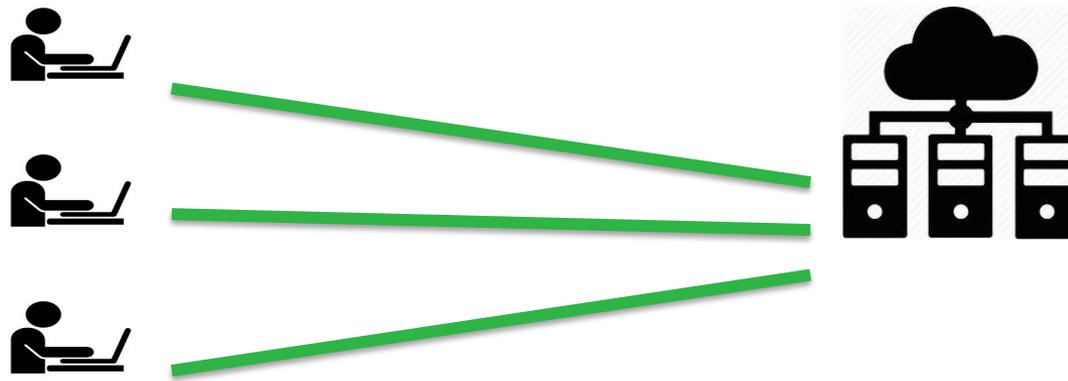
*Comment travailler sur le serveur ?*



# Comment travailler sur le serveur ?



- En se connectant sur un serveur linux distant de son ordinateur via le **protocole ssh**



## HPC South Green

- itrop (IRD)

[bioinfo-inter.ird.fr](http://bioinfo-inter.ird.fr)

- HPC AGAP (CIRAD)

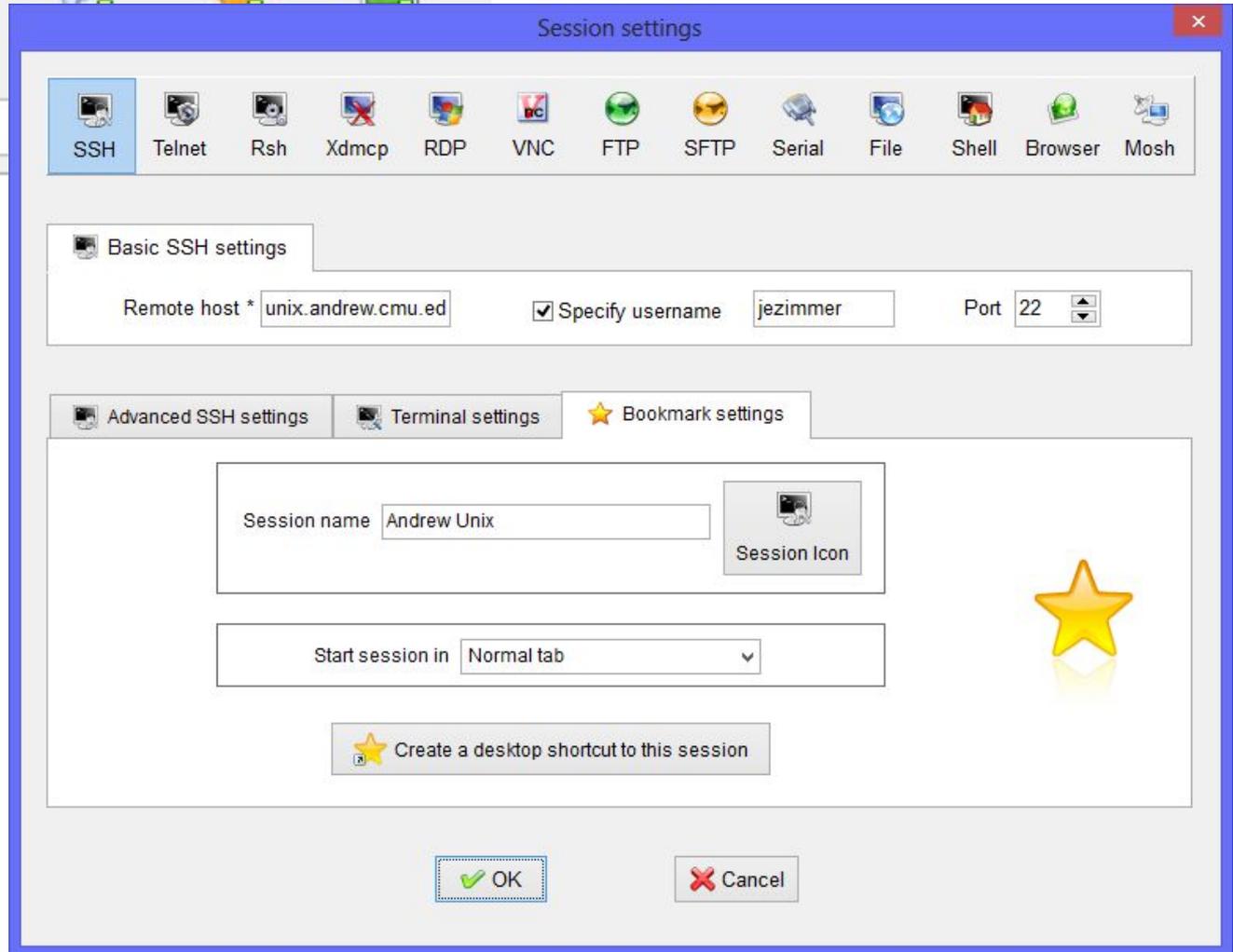
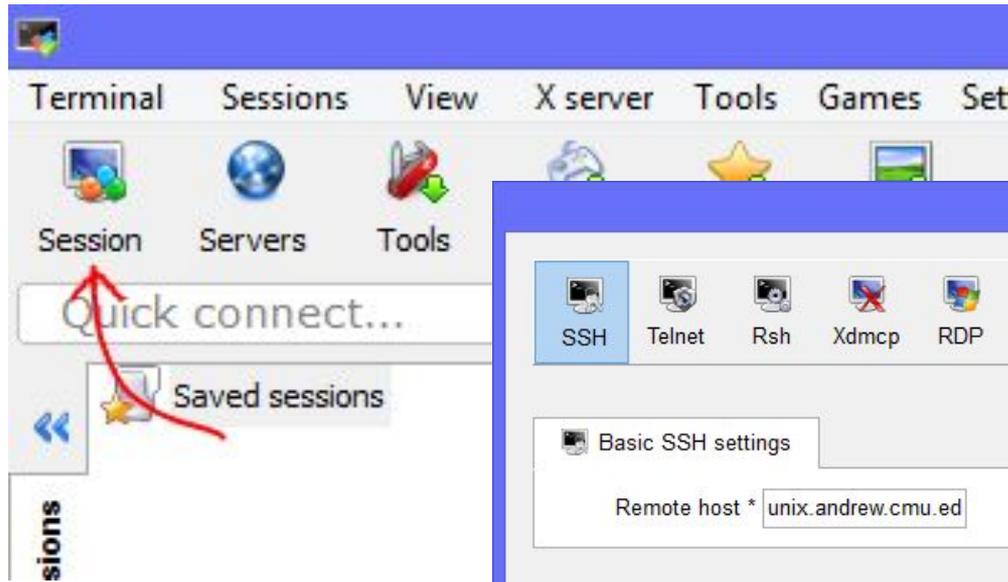
[cc2-login.cirad.fr](http://cc2-login.cirad.fr)



PuTTY



# Comment travailler sur le serveur ?





# *Comment travailler sur le serveur ?*

```
tranchan — CLUSTER — ssh bioinfo-inter.ird.fr -ltranchant — 13
Last login: Sat Mar 16 11:48:06 on ttys002
MacBook-Pro-de-Christine:~ tranchan$ ssh bioinfo-inter.ird.fr -ltranchant
Warning: Permanently added the ECDSA host key for IP address '64:ff9b::5bcb:2296'
Enter passphrase for key '/Users/tranchan/.ssh/id_rsa':
```



# Practice

putty,  
terminal, ssh

2

Go to [Practice 2](#) on our github



# Premiers Pas sur Linux

Commandes relatives à l'arborescence de  
fichiers

Toujours présent sur le terminal, juste avant de taper la commande

**Prompt**

```
[tranchant@node6 data]$ █
```

Toujours présent sur le terminal, juste avant de taper la commande

**Prompt**

```
[tranchant@node6 data]$
```

Nom  
utilisateur

Nom  
serveur

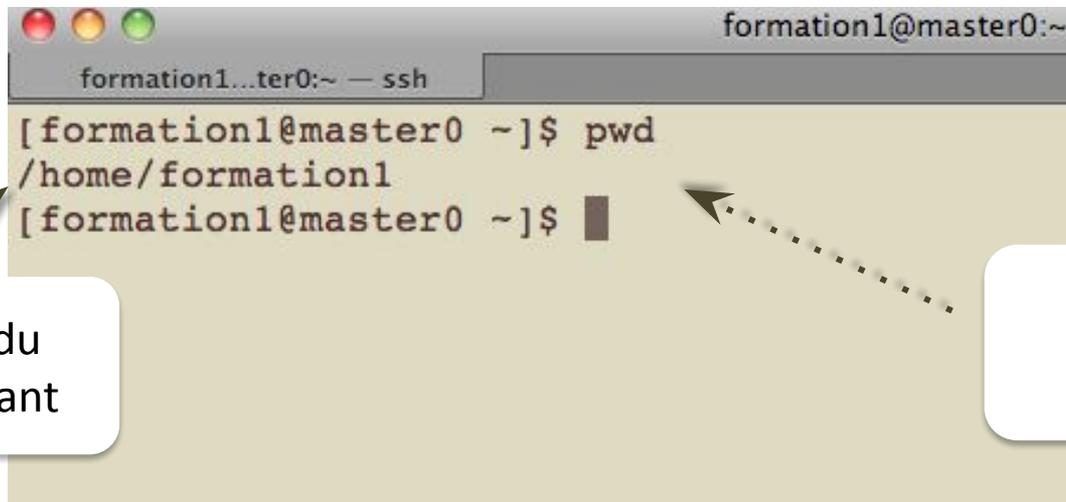
Répertoire  
courant

*commande [ -options ] [ arguments ]*

**pwd**

*Present Work Directory*

*affiche nom du répertoire courant (et son chemin complet)*



```
formation1@master0:~  
formation1...ter0:~ — ssh  
[formation1@master0 ~]$ pwd  
/home/formation1  
[formation1@master0 ~]$ █
```

Affiche le nom du répertoire courant

Commande sans options et sans argument

**ls**  
list

*liste fichiers & répertoires présents dans le répertoire courant*

```
formation1@master0:~ — ssh — 97x37  
formation1@master0:~ — ssh  
[formation1@master0 ~]$ ls  
data  scripts
```

Liste des fichiers du répertoire courant (par défaut)

Commande sans options et sans argument

**ls -l**  
*list long*

*liste les fichiers avec des informations pour chaque fichier/répertoire*

Commande avec l'option l et le nom d'un répertoire en argument

```
[formation1@master0 ~]$ ls -l /home/
total 312
drwx-----  6 abate      sat          4096 12 mars   2012 abate
drwx-----  5 adam        ggr          4096 23 mars   2012 adam
drwx----- 31 admin      admin        4096  3 août   11:35 admin
drwx-----  9 alizon     ete          4096 21 août   14:23 alizon
drwx----- 12 alvaro-wis effecteurs   4096 17 juin   16:19 alvaro-wis
drwx-----  4 auguy     rhizogenesis 4096  2 mars   2012 auguy
drwx-----  5 ayouba    team1        4096 13 avril  2012 ayouba
drwx-----  5 beule     bdp          4096  8 oct.   17:49 beule
drwx-----  9 bouniol   ggr          4096  2 oct.   15:00 bouniol
drwx----- 10 castillo  bdp          4096 10 oct.   15:55 castillo
```

liste détaillée des fichiers

## Comment obtenir de l'aide sur une commande?

- avec l'option *--help* ou *-h*      *ls --help*      *blastn -h*
- avec la commande *man*      *man ls*
- avec la commande *whatis*      *whatis ls*

```
tranchan — tranchant@master0:~ — ssh bioinfo-master.ird.fr -ltranchant — 93x22
...nt@master0:~ — ssh bioinfo-master.ird.fr -ltranchant  ~ — tranchant@master0:~ — -bash  ...  ~ — CLUSTER — -bash
[tranchant@master0 ~]$ whatis pwd
pwd (1)          - Afficher le nom du répertoire de travail en cours
pwd (n)          - Return the absolute path of the current working directory
pwd (lp)         - return working directory name
[tranchant@master0 ~]$
```

speaker notes

## Arborescence linux

**pwd**

Affiche le chemin absolu

**ls**

Liste tous les fichiers/répertoires

**ls -l**

Affiche toutes les informations sur les fichiers



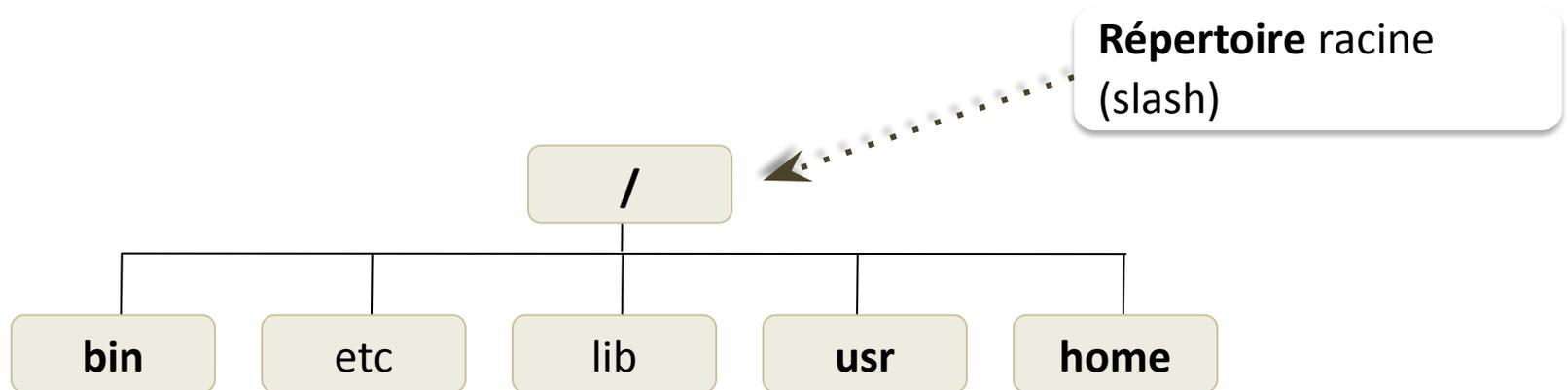
# Practice

prompt, pwd

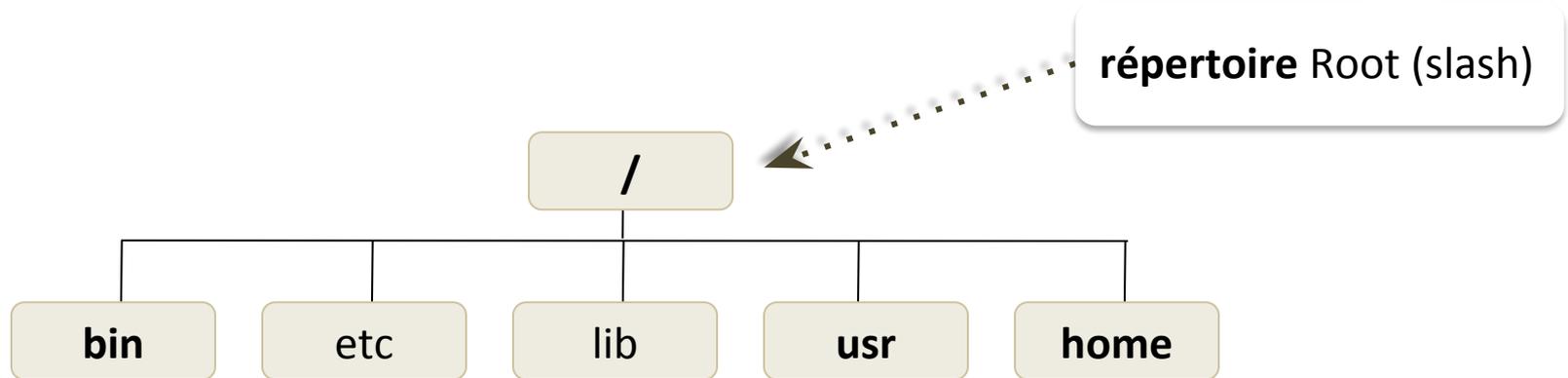
3

Go to [Practice 3](#) on our github

- Arborescence = système de fichiers
- Répertoire racine appelé “/”



## Principaux répertoires



<b>/bin</b>	<b>principales commandes système</b>
<b>/usr, /opt</b>	<b>Applications et librairies ex : /usr/local</b>
<b>/usr/bin</b>	<b>Autres commandes</b>
<b>/home</b>	<b>répertoire des utilisateurs (1 par utilisateur, name=login)</b>

**Chemin (path)** : chemin d'accès à un fichier/répertoire

**Chemin (path)** : chemin d'accès à un fichier/répertoire

absolu

- chemin complet du fichier en partant du répertoire racine /

**Chemin** (path) : chemin d'accès à un fichier/répertoire

## absolu

- chemin complet du fichier en partant du répertoire racine /
- *commence toujours par /*
- **Toujours correct peu importe où on l'on travaille**

**Chemin** (path) : chemin d'accès à un fichier/répertoire

## absolu

- chemin complet du fichier en partant du répertoire racine /
- commence toujours par /
- Toujours correct peut importe où on l'on travaille

## relatif

- chemin défini par rapport à où l'on est dans l'arborescence

**Chemin** (path) : chemin d'accès à un fichier/répertoire

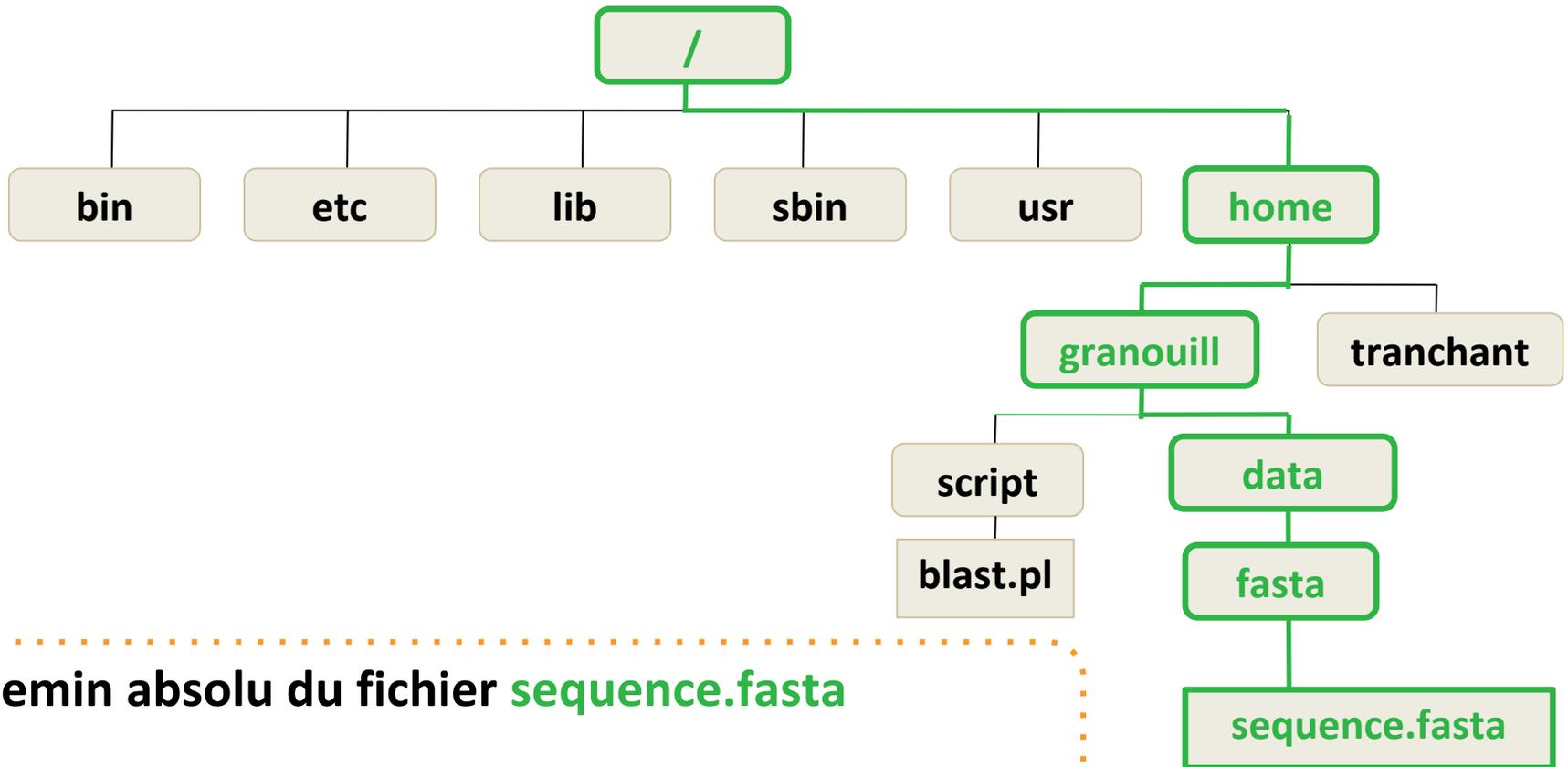
## absolu

- chemin complet du fichier en partant du répertoire racine /
- commence toujours par /
- Toujours correct peut importe où on l'on travaille

## relatif

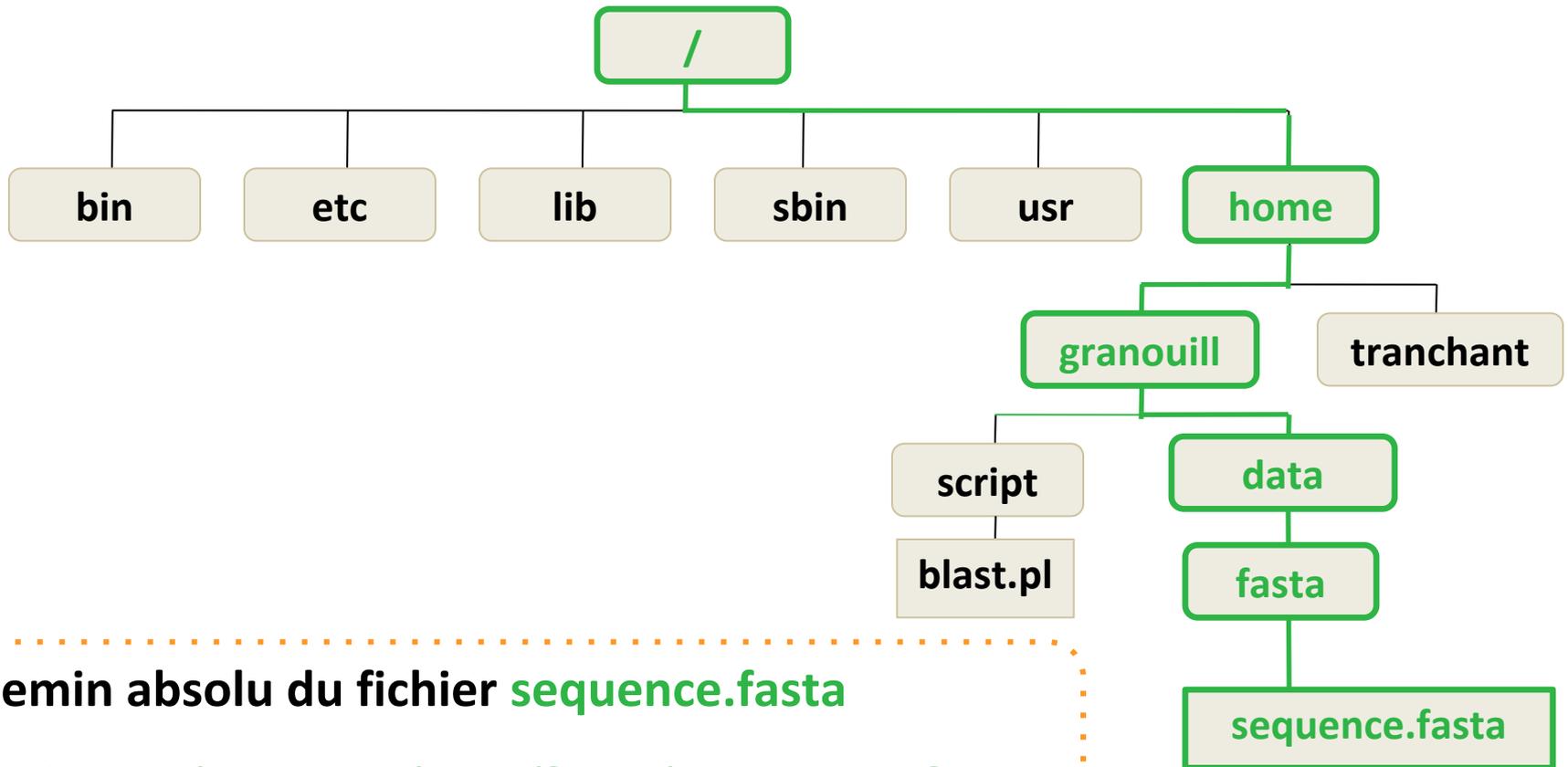
- chemin défini par rapport à où l'on est dans l'arborescence
- ***Ne commence jamais par /***
- **Change selon où l'on travaille**

- Commence toujours par / (le répertoire racine)
- **Toujours correct peu importe où l'on travaille**



Chemin absolu du fichier **sequence.fasta**

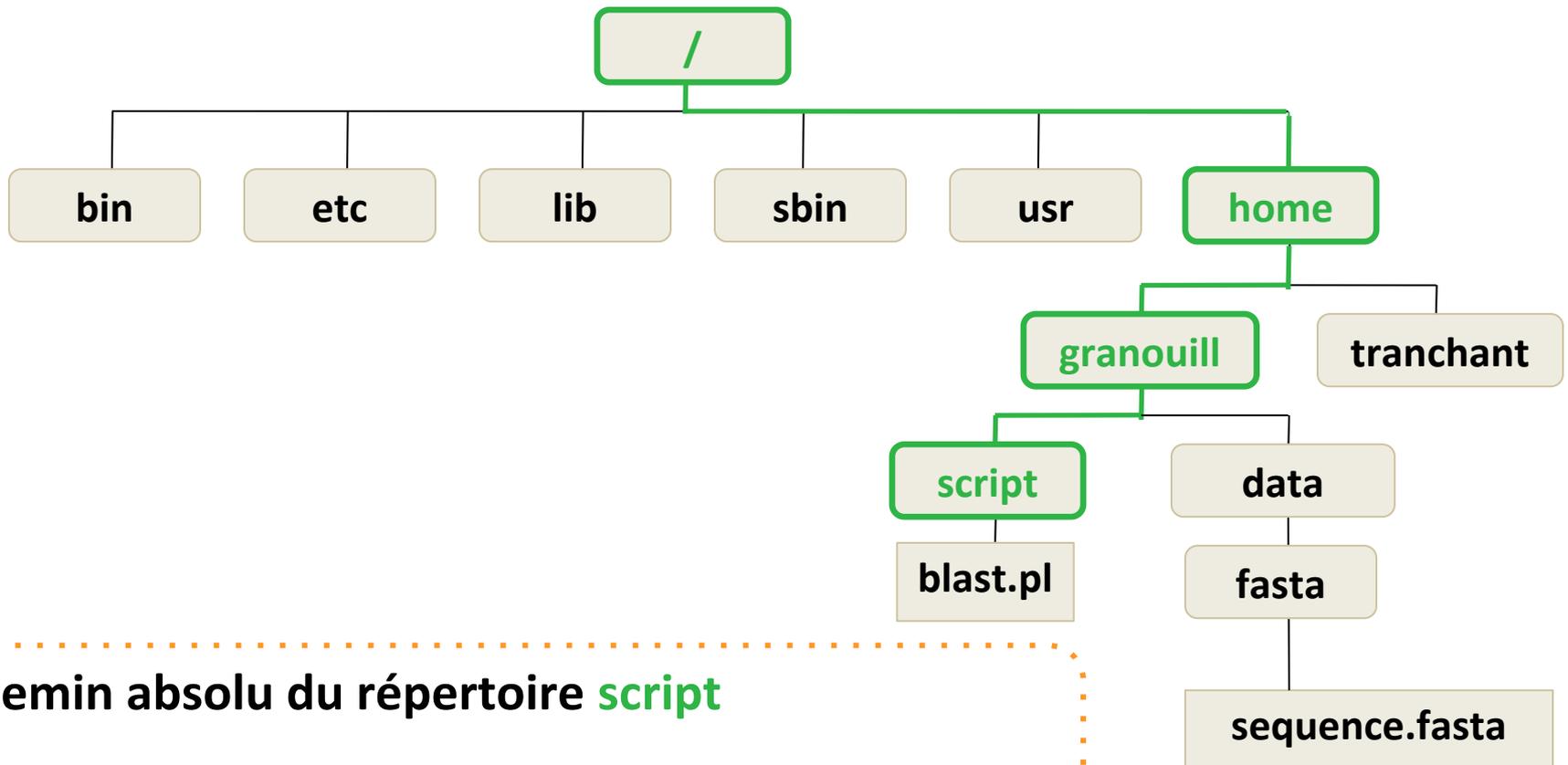
- Commence toujours par / (le répertoire racine)
- **Toujours correct peu importe où l'on travaille**



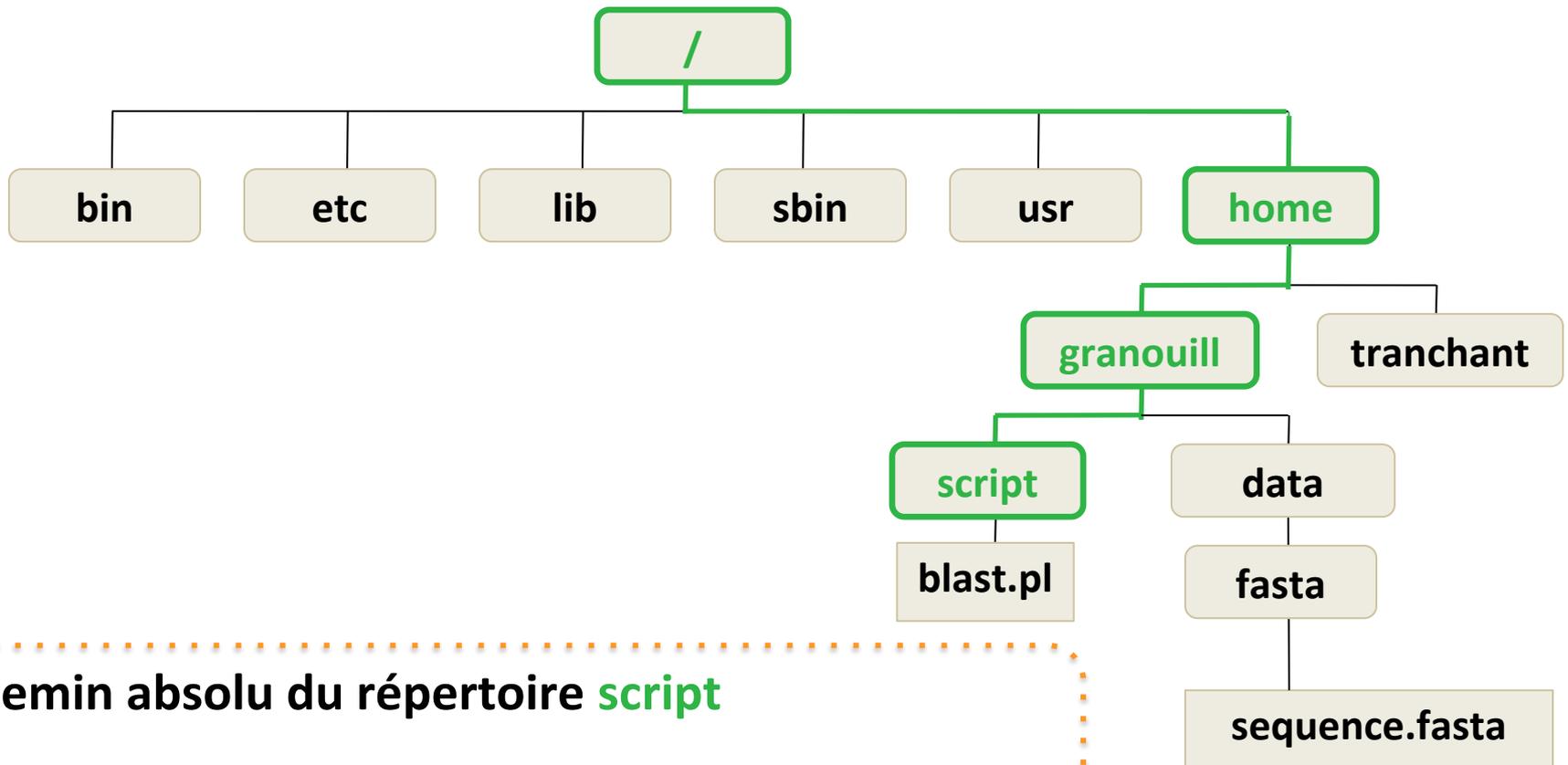
Chemin absolu du fichier **sequence.fasta**

`/home/granouill/data/fasta/sequence.fasta`

- Commence toujours par / (le répertoire racine)
- **Toujours correct peu importe où l'on travaille**



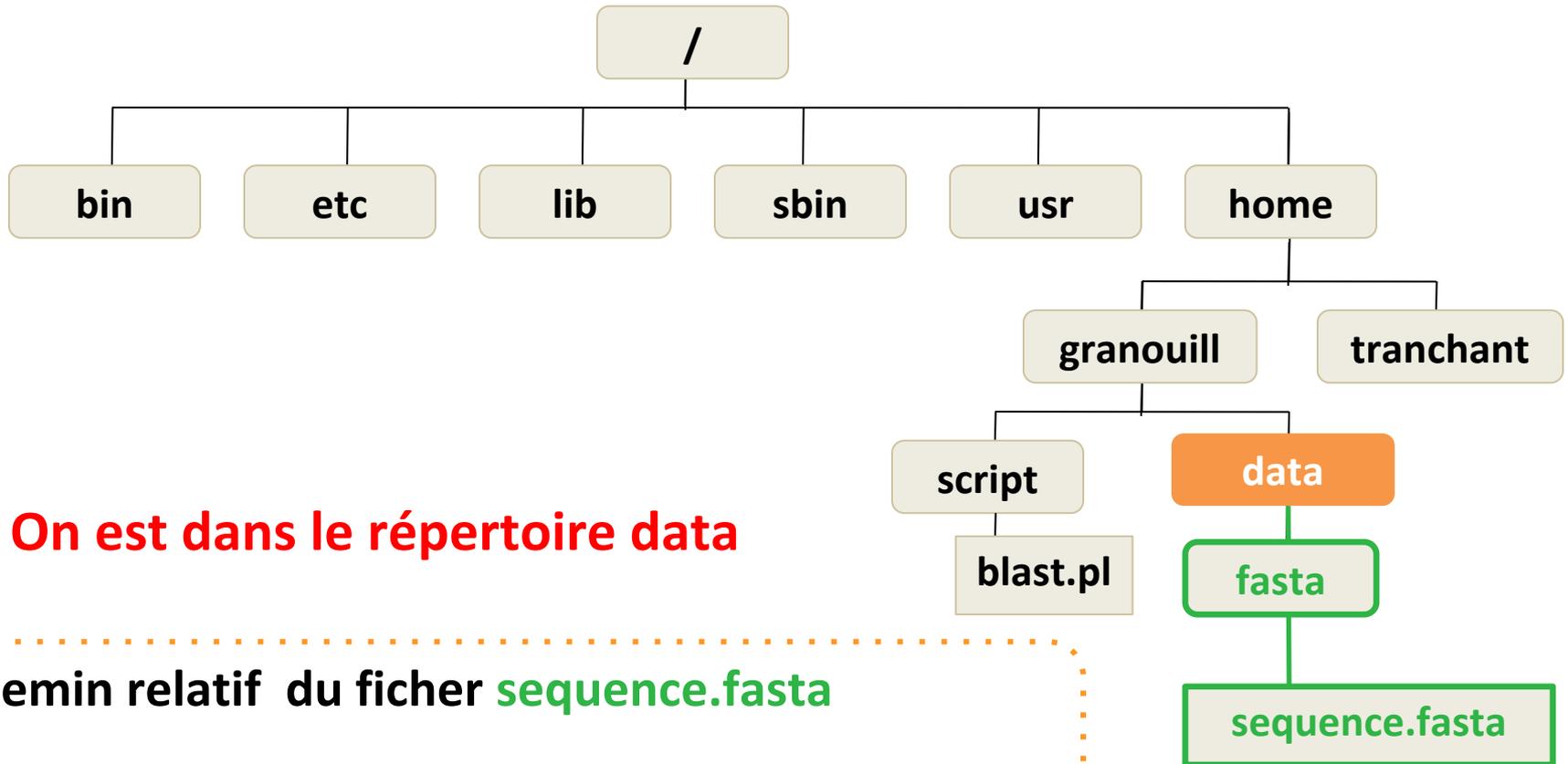
- Commence toujours par / (le répertoire racine)
- **Toujours correct peu importe où l'on travaille**



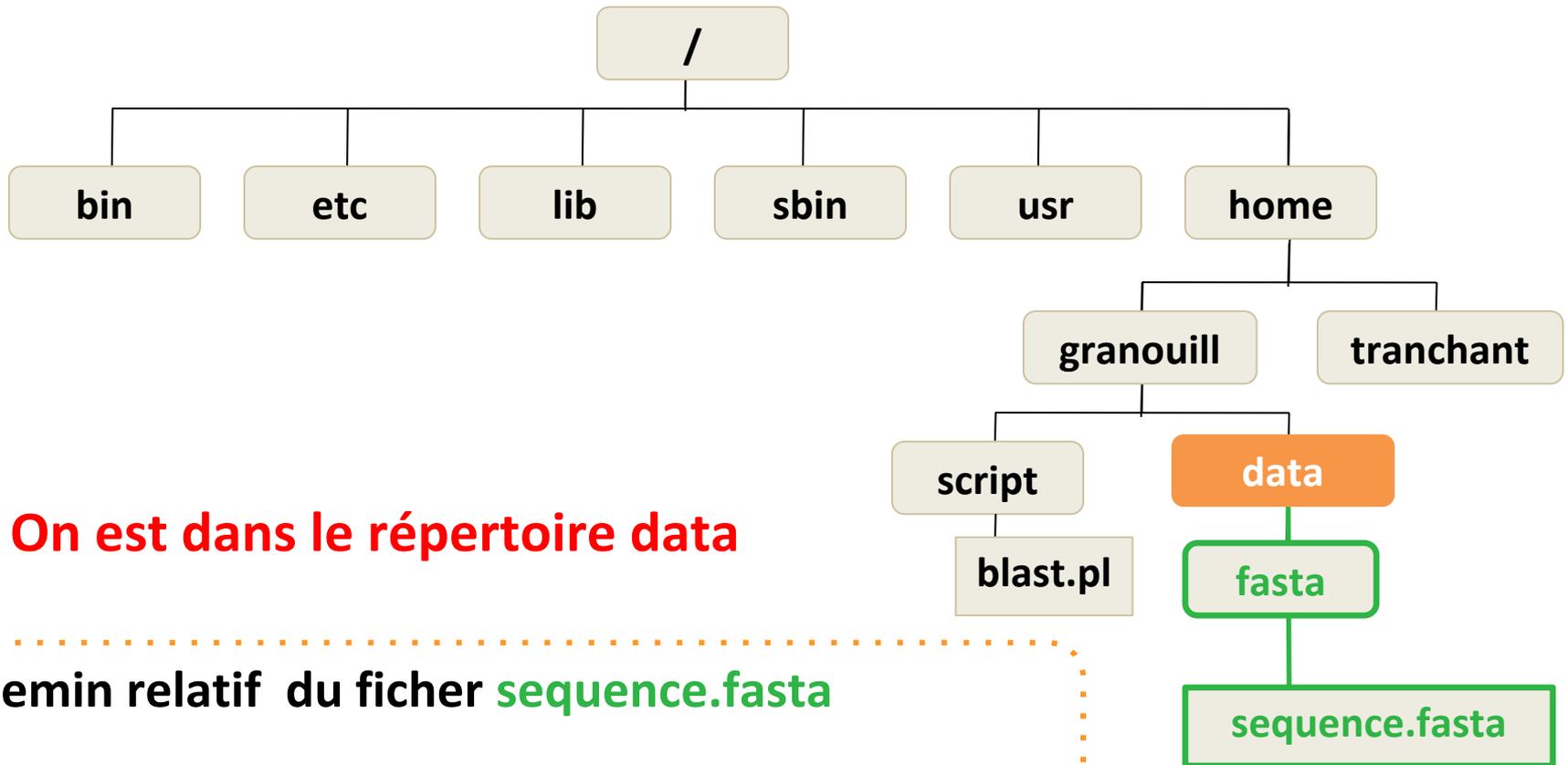
Chemin absolu du répertoire **script**

`/home/granouill/data/script`

- Défini par rapport où on est dans l'arborescence
- **Ne commence jamais par /**



- Défini par rapport où on est dans l'arborescence
- **Ne commence jamais par /**

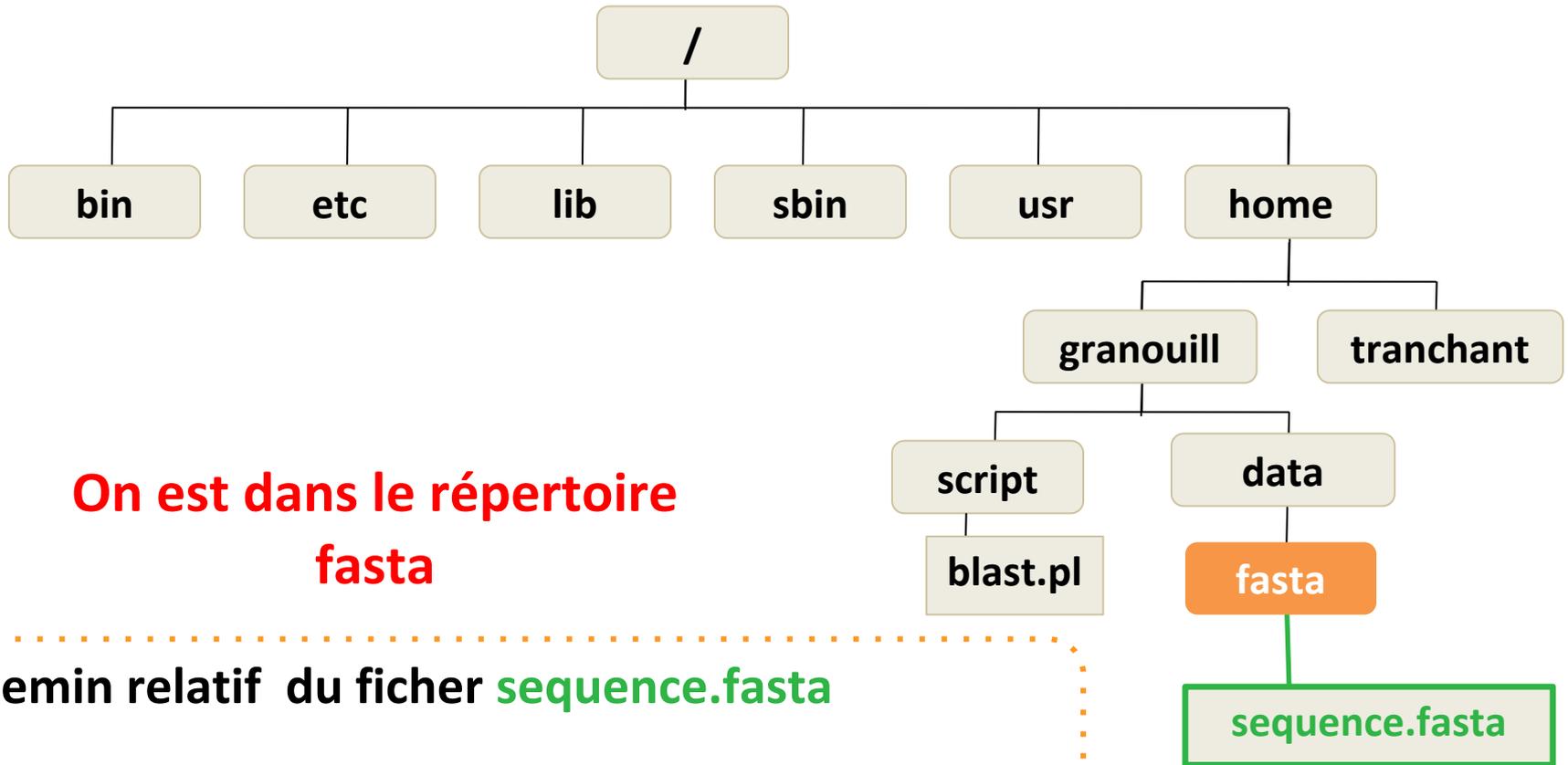


**On est dans le répertoire data**

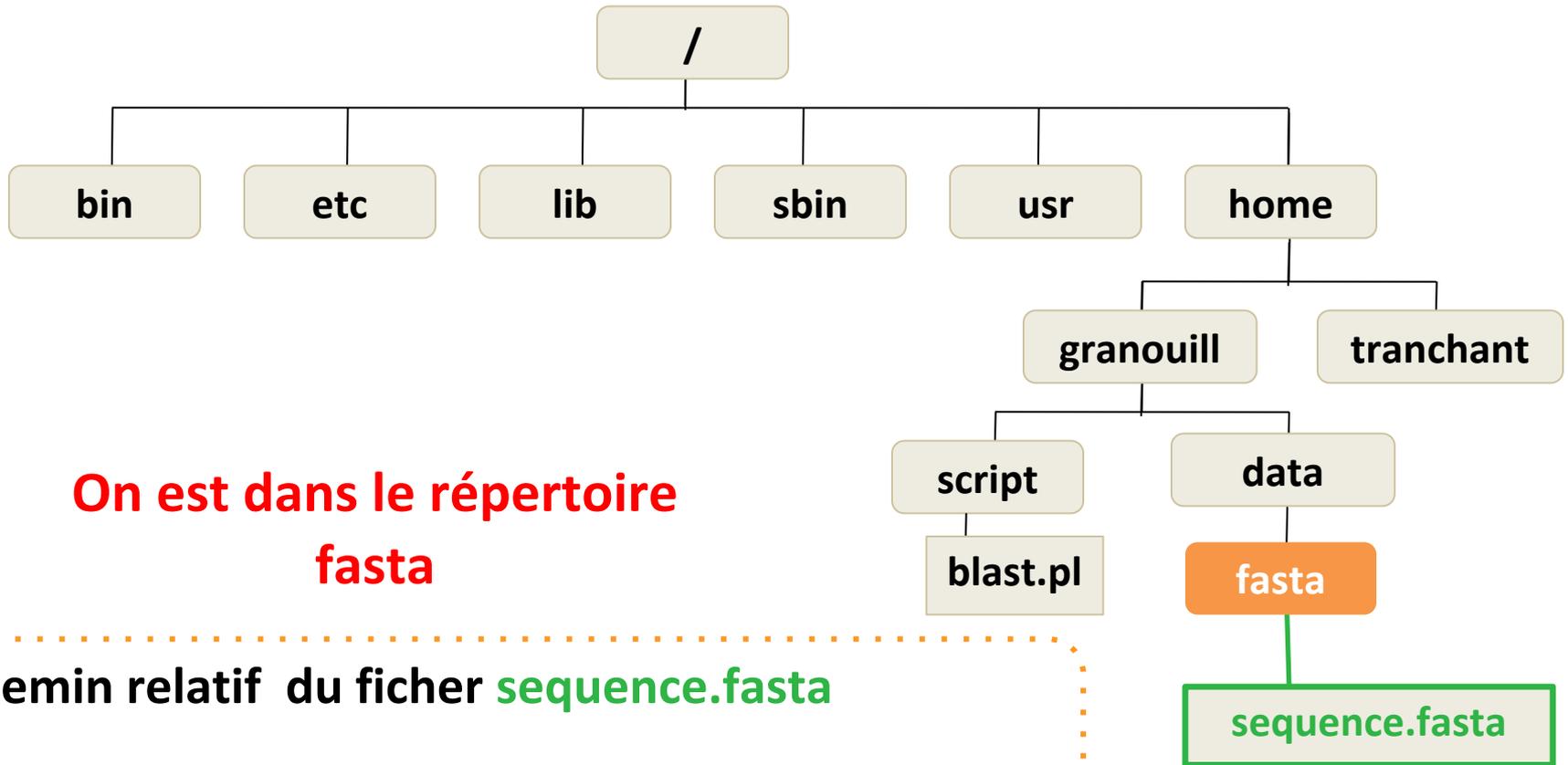
Chemin relatif du fichier `sequence.fasta`

`fasta/sequence.fasta`

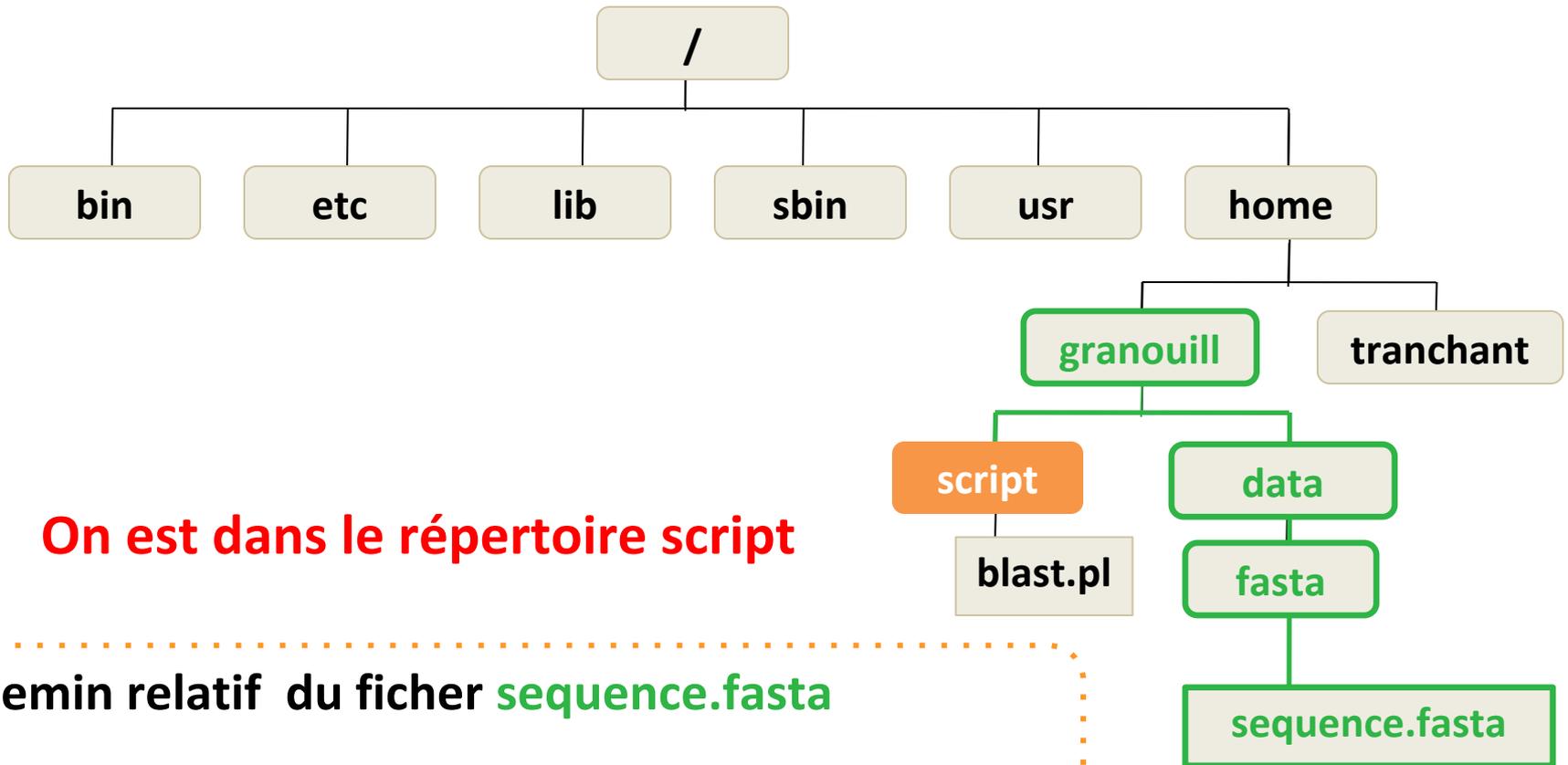
- Défini par rapport où on est dans l'arborescence
- **Ne commence jamais par /**



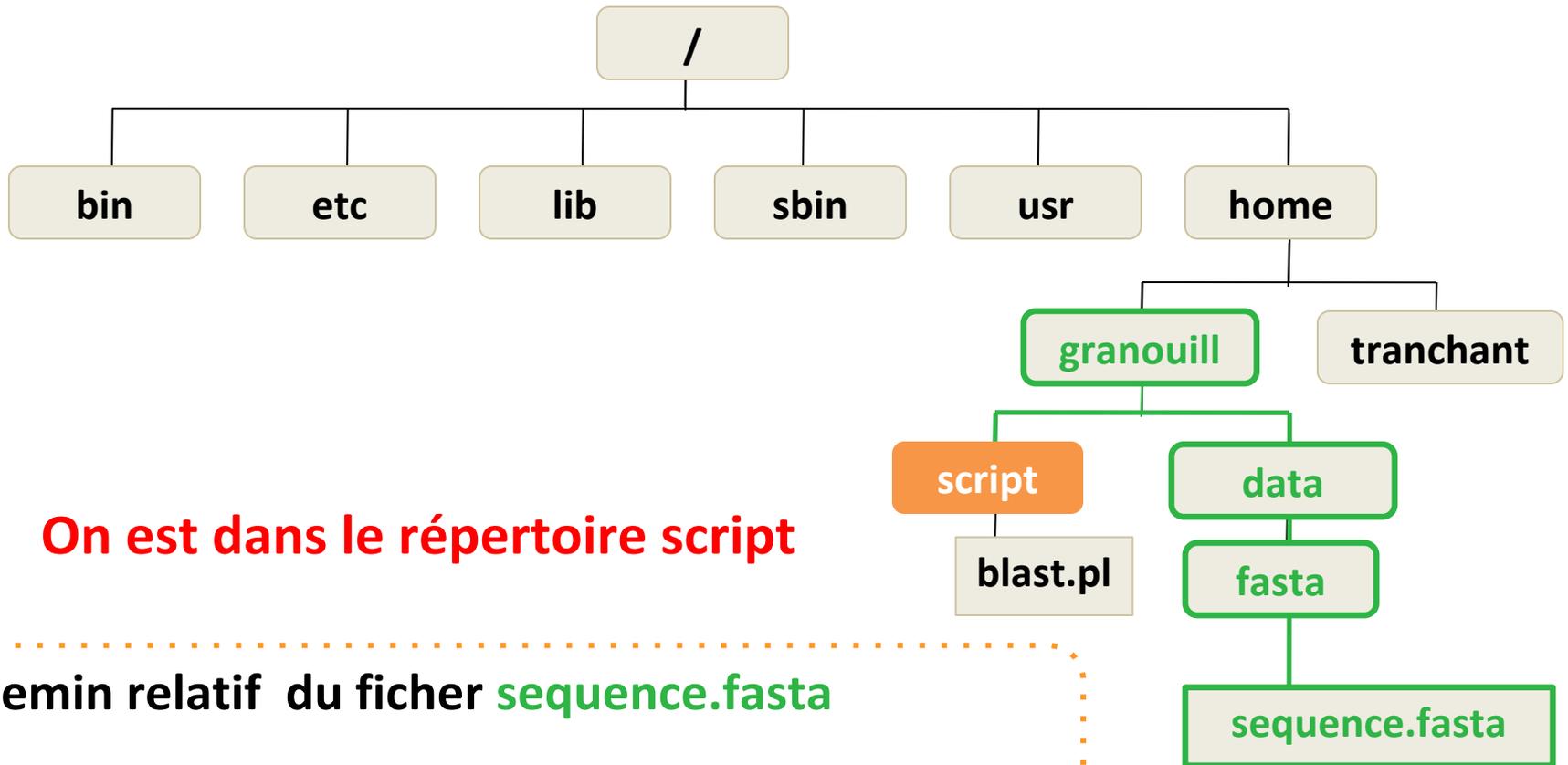
- Défini par rapport où on est dans l'arborescence
- **Ne commence jamais par /**



- Défini par rapport où on est dans l'arborescence
- **Ne commence jamais par /**



- Défini par rapport où on est dans l'arborescence
- **Ne commence jamais par /**



**On est dans le répertoire script**

Chemin relatif du fichier **sequence.fasta**

**../data/fasta/sequence.fasta**



# Practice

Is

4

Go to [Practice 4](#) on our github

## Interagir avec les processus

- <Ctrl> + C** Arrêter le processus en cours sous le terminal
- <Ctrl> + Z**

## Tab completion

- <Tab>** Complète automatiquement le nom d'un fichier/ répertoire qui est en cours de saisie (choix unique)
- <Tab><Tab>** Affiche la liste des différentes possibilités si le choix n'est pas unique

## Interagir avec l'historique de commandes

**Flèche bas/haut** - Afficher la commande précédente/suivante  
- Presser plusieurs fois pour naviguer dans l'historique

**<Ctrl> + R** - Afficher la dernière commande qui contient les caractères saisis.  
- Presser les touches et commencer à taper la commande recherchée

- Linux est sensible à la casse

Sequence.fasta **≠** SEQUENCE.fasta **≠** sequence.fasta

- Utiliser uniquement des lettres, chiffres et caractères \_ . –

- PAS d'espaces, accents et caractères spéciaux

& ~ # " ' { ( [ | ` \ ^ @ ) ] } \$ \* % ! / ; , ?

- Le suffixe des noms de fichiers (.txt, .fasta, .fa, .fq etc.) est optionnel



# Travailler avec les “jokers”

Métacaractères : \*, [ ]

C'est un caractère qui peut être utilisé comme substitut de plusieurs caractères ou classes de caractère dans une recherche

Peut être utilisé avec de nombreuses commandes linux

\*

N'importe quelle chaîne de caractère



KYVF-01.R1.fastq

KYVF-02.R1.fastq

KYVF.sam

ZO16.fastq

KYVF-01.R2.fastq

KYVF-02.R2.fastq

KYVF.bam

ZO16.bam

```
ls *fastq
```

\*

N'importe quelle chaîne de caractère



KYVF-01.R1.fastq

KYVF-02.R1.fastq

KYVF.sam

ZO16.fastq

KYVF-01.R2.fastq

KYVF-02.R2.fastq

KYVF.bam

ZO16.bam

```
ls *fastq
```

KYVF-01.R1.fastq

KYVF-02.R1.fastq

ZO16.fastq

KYVF-01.R2.fastq

KYVF-02.R2.fastq

\*

N'importe quelle chaîne de caractère



KYVF-01.R1.fastq

KYVF-02.R1.fastq

KYVF.sam

ZO16.fastq

KYVF-01.R2.fastq

KYVF-02.R2.fastq

KYVF.bam

ZO16.bam

```
ls KYVF*fastq
```

\*

N'importe quelle chaîne de caractère



KYVF-01.R1.fastq

KYVF-02.R1.fastq

KYVF.sam

ZO16.fastq

KYVF-01.R2.fastq

KYVF-02.R2.fastq

KYVF.bam

ZO16.bam

```
ls KYVF*fastq
```

```
KYVF-01.R1.fastq
```

```
KYVF-02.R1.fastq
```

```
KYVF-01.R2.fastq
```

```
KYVF-02.R2.fastq
```

- \* N'importe quelle chaîne de caractère
- [ensemble]** Tous les caractères dans l'ensemble
- [!ensemble]** Tous les caractères pas dans l'ensemble



KYVF-01.R1.fastq	KYVF-02.R1.fastq	KYVF.sam	ZO16.fastq
KYVF-01.R2.fastq	KYVF-02.R2.fastq	KYVF.bam	ZO16.bam

```
ls *. [sb]am
```

- \* N'importe quelle chaîne de caractère
- [ensemble] Tous les caractères dans l'ensemble
- [!ensemble] Tous les caractères pas dans l'ensemble



KYVF-01.R1.fastq	KYVF-02.R1.fastq	KYVF.sam	ZO16.fastq
KYVF-01.R2.fastq	KYVF-02.R2.fastq	KYVF.bam	ZO16.bam

```
ls *. [sb]am
```

```
KYVF.sam ZO16.bam
```

```
KYVF.bam
```

```
= ls *. [!f]*
```



# Practice

ls, \*

5

Go to [Practice 5](#) on our github



# Commandes relative à l'arborescence de fichiers

commande cd

**cd**

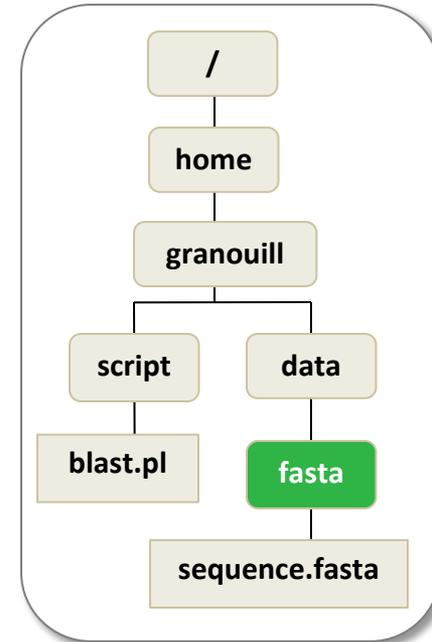
*pour se déplacer dans l'arborescence*

*Change Directory*

*cd nom\_repertoire (chemin absolu ou relatif)*

`cd nom_repertoire(chemin absolu ou relatif)`

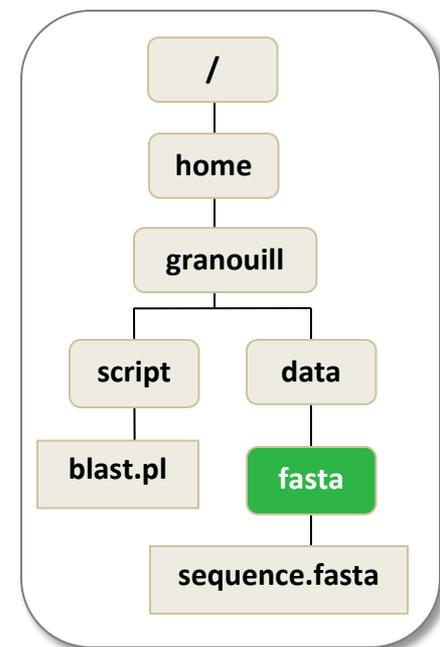
**Chemin absolu :**



`cd nom_repertoire(chemin absolu ou relatif)`

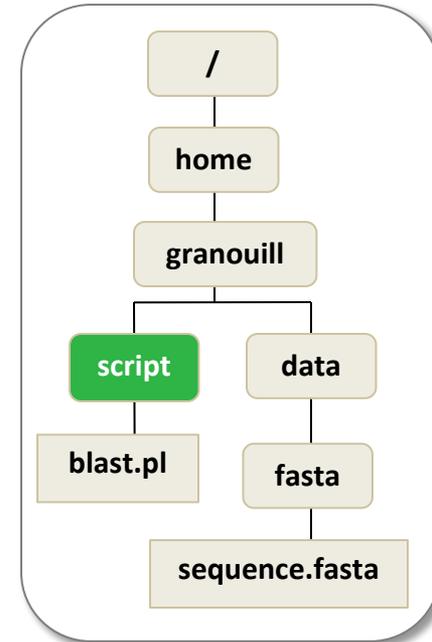
**Chemin absolu :**

`cd /home/granouill/data/fasta`



`cd nom_repertoire(chemin absolu ou relatif)`

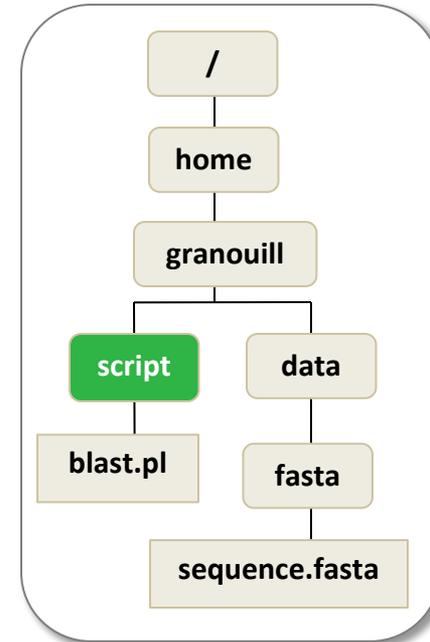
**Chemin absolu :**



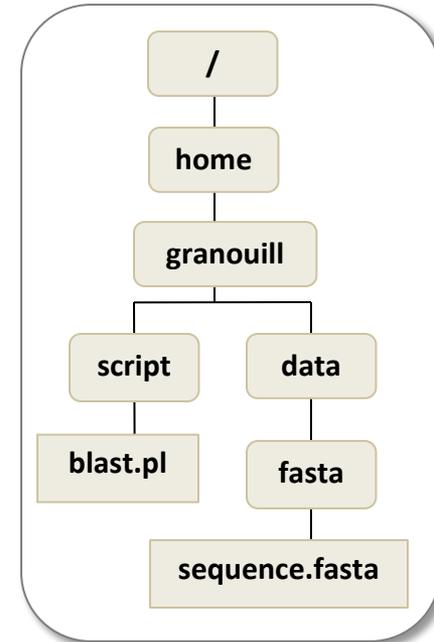
`cd nom_repertoire(chemin absolu ou relatif)`

**Chemin absolu :**

`cd /home/granouill/script/`



`cd nom_repertoire(chemin absolu ou relatif)`



## Chemin relatif :

Commande	Se déplace dans	
<code>cd directory_name</code>	<code>directory_name</code>	
<code>cd</code>	<i>home directory</i>	Retour au répertoire de l'utilisateur courant home
<code>cd ..</code>	le répertoire parent	Remonte d'1 répertoire
<code>cd ../..</code>	le répertoire parent du répertoire parent	Remonte de 2 répertoires

## Arborescence linux

**pwd**

Affiche le chemin absolu

**ls**

Liste tous les fichiers/répertoires

**cd rep\_name**

Se déplace dans rep\_name

**mkdir rep\_name**

Crée un répertoire

**rm nom\_fichier**

Supprime un fichier

**cp fichier\_source repertoire\_cible**

Copie un fichier dans un répertoire

**cp fichier\_source nouveau\_fichier**

Copie un fichier sous un nouveau nom

**cp fichier\_source repertoire\_cible/nouveau\_fichier**

Mix des 2



# Practice

`pwd, cd, ls`

6

*Go to [Practice 6](#) on our github*



# Practice

`mkdir, mv, ls, cp, rm`

7

Go to [Practice 7](#) on our github



**Des commandes pour éditer  
des fichiers et les manipuler**

**cat nom\_fichier**

*Affiche le contenu d'un fichier sans  
pagination  
(à éviter s'il est volumineux!!)*

```
MacBook-Pro-de-Christine:Data tranchan$ cat Data/Fasta/sequence.fasta  
>Gxbjbsjxbjs  
CCACCCCTCTTACAGTCTTCACCAAATGTCCTTTAAACTCCACCTAAAGTATCCAAAGA  
CTCGAGAAATGCTGTGCCACAACCAGCTTTTGAGTCATCCATGACCGTTGATCTTCCTTT  
GCCCCAGAGTGGGGCCTAGCACCATCTAGCTACTACTTGCCTTTCATACCCATCATTGG  
GATACCCTGAATACCTATCTTATAAGTTCCATATGGCTTATATTTCTAAGTAAGAGATGC  
ACTTAGTAAGTGCATGTCGTCCTGACTTGTTTATACTCTAATGTATGATATTTATATCCC  
TATAATATAGTGTTACTAATATATGTTTGGTATTGTGTAGACTCCATTGTACCATGGTGT  
GCTAATTAGAAATAACATGCCAGCTTTGCTATTGTGGTTTGCAAGTAAAGTAAAAAAAAA  
MacBook-Pro-de-Christine:Data tranchan$
```

## less nom\_fichier

*Affiche le contenu d'un fichier*

<space> pour défiler

q pour quitter

/ suivi par le mot recherché

*less Data/Fasta/EST-68566-Coffeacaneophora.fasta*

```
>gi|315911382|gb|GT649553.1|GT649553 CC00-XX-SH3-053-F02-EM.F Coffea canephora SH3 Coffea canephora cDNA clone CC00-X
X-SH3-053-F02-EM, mRNA sequence
AATATTTTGGGTGTCACCGTTCACCGCTAATCTTTCTGCACAAAGTAAGAGAAGGGTAACAAGTAAAAGGG
AGCGAGACTATGGCTGCCACATCTGCTGCAGTACTAAATGGATTGGGCTCTCCCTTCTTGACTGGTGGAA
GGAAAAGTCAGGCCCTGCTGGCTTCACCAGCTGCAGCTAGAGTCGGTGGTGTGCTGCTGTTGCTCCAAA
GAGATTAGTTGTGGTAGCTGCTCGTCCCTCCCAAGAAGTCATGGATTCCCTGGTGTAGAACTGGTCCAGGC
TTCCTTGACCCCGAGTATCTCGATGGCTCGCTCCCTGGTACTACGGTTTTGATCCACTTGGTCTTGGCA
AGGATCCAGCATTCTTGAATGGTATAGAGAAGCTGAGCTCATTCATGGCAGATGGGCAATGACTGCAGT
TGTGGAACTTTGTTGGCCAGGCATGGAGTGGTATCCCATGGTTTCAAGCTGGCGCTGACCCCTGGTGCC
ATTGCGCCATTCTCCTTTGGTTCGCTCCTCGGTACTCAACTCCTGCTCA

>gi|315911381|gb|GT649552.1|GT649552 CC00-XX-SH3-077-D04-EM.F Coffea canephora SH3 Coffea canephora cDNA clone CC00-X
X-SH3-077-D04-EM, mRNA sequence
ATTTCATCGTATCTTGCCCTCTTTCTTTTCATCTCTCCTCTTCTTTTGTATCAGCATCAGAGGCACCGGCAGC
TTTTAAAAGCTTTATCACACACACACACATTTCTTTGTAGTCAGTCAGTGGCAACAATTACTGAGTACT
ACTAGTTACTTATCGAACAATTCGATTAGCTTTGCTAGGAGGTACCGTTTCGGTTGGTTGACTTGTGTG
AGTATAAATATTGATGGCTGGGATGCTGTGCGGAGTTAACATTTGGAGGCGAGACTGAAACAGCTAAACCA
GTCGAGCCTAATTTCTCAGTCCGCTAGGCCGAGGAGGATGGAATTCATCATTTCATATTGTAGCTTCCGG
Data/Fasta/EST-68566-Coffeacaneophora.fasta
```

**head** Affiche les premières lignes d'un fichier  
(n=10, 10 lignes par défaut) `head -n 20  
script.pl`

**tail** affiche les dernières lignes d'un fichier  
(n=10 par défaut) `tail -n 5 script.pl`

**wc** Compter le nombre de lignes, mots ou  
caractères d'un fichier `wc script.pl  
wc -l script.pl`



# Practice

Is

8

Go to [Practice 8](#) on our github

**grep**

*pour rechercher un motif dans un fichier*

```
grep [options] motif [file1, ... ]
```

**Pour rechercher un motif, le mettre entre simple/double quote  
=> ex : 'gene' ou "gene"**

**grep**

*pour rechercher un motif dans un fichier*

```
grep [options] motif [file1, ... ]
```

Option	Description
-i	Recherche le motif sans tenir compte de la casse
-c	Compte le nombre de lignes dans lesquelles le motif a été trouvées
-v	Affiche seulement les lignes sans le motif
-l	Affiche uniquement les noms de fichiers dans lesquels le motif a été trouvé

**grep**

*pour rechercher un motif dans un fichier*

```
grep [options] motif [file1, ... ]
```

Option	Description
-c	Compte le nombre de lignes dans lesquelles le motif a été trouvées
-l	Affiche uniquement les noms de fichiers dans lesquels le motif a été trouvé
-i	Recherche le motif sans tenir compte de la casse
-v	Affiche seulement les lignes sans le motif



# Practice

9

Go to [Practice 9](#) on our github

**cut**

Sélection de colonnes/champs d'un fichier

```
cut -d separateurColonne -f fieldNumber fileName
```

```
cut -d ":" -f1,5 /etc/passwd
```

Sélection des champs 1 et 5 dans un fichier tabulé  
dont les colonnes sont séparées par un :

## *sort*

Trier une ou plusieurs colonnes d'un fichier tabulé

```
sort -k2 fileName
```

sur la colonne 2 (tri alphanumérique)

```
sort -k2r fileName
```

sur la colonne 2 (décroissant)

```
sort -k2g -k1r fileName
```

sur les colonnes 2 (numérique) et 1 (décroissant)

```
sort -t: -k3g fileName
```

sur la colonne 3 (numérique), le délimiteur de  
colonne étant :



# Practice

10

*Go to [Practice 10](#) on our github*



# Les entrées / sorties

**pour sauvegarder la sortie d'une  
commande dans un fichier**

La sortie d'une commande peut être écrite

dans un fichier avec les caractères :

>

>>

```
cut -d: -f1 /etc/passwd > userName.txt
```

```
cut -d: -f1 /etc/passwd > userName.txt
```

## Redirection

## Action

Command **>** file

- si le fichier n'existe pas : il sera créé
- si le fichier existe : efface le contenu

Command **>>** file

- si le fichier n'existe pas : il sera créé
- si le fichier existe : écrit à la fin du fichier



# Practice

2

Go to [Practice 11](#) on our github

- La sortie d'une 1ère commande peut être envoyée comme input d'une 2ème commande
- Pour connecter/combiner plusieurs commandes ensemble dans la même ligne de commande (sans fichier intermédiaire)
- Pipelines= *workflow*

*cmd1 | cmd2 | cmd3*

```
cut -d: -f1 /etc/passwd
```

```
Root
```

```
troot
```

```
iroot
```

```
ctroot
```

```
//
```

```
cut -d: -f1 /etc/passwd
```

```
Root  
troot  
iroot  
ctroot  
//
```

```
cut -d: -f1 /etc/passwd | sort
```

```
abate  
adm  
adroot  
ais  
#albar  
alvaro-wis  
anthony  
apache
```

```
cut -d: -f1 /etc/passwd
```

```
Root  
troot  
iroot  
ctroot  
//
```

```
cut -d: -f1 /etc/passwd | sort
```

```
abate  
adm  
adroot  
ais  
#albar  
alvaro-wis  
anthony  
apache
```

```
cut -d: -f1 /etc/passwd | sort | head
```

```
cut -d: -f1 /etc/passwd
```

```
Root  
troot  
iroot  
ctroot  
//
```

```
cut -d: -f1 /etc/passwd | sort
```

```
abate  
adm  
adroot  
ais  
#albar  
alvaro-wis  
anthony  
apache
```

```
cut -d: -f1 /etc/passwd | sort > /etc/passwd.sort
```



# Practice

3

Go to [Practice 12](#) on our github

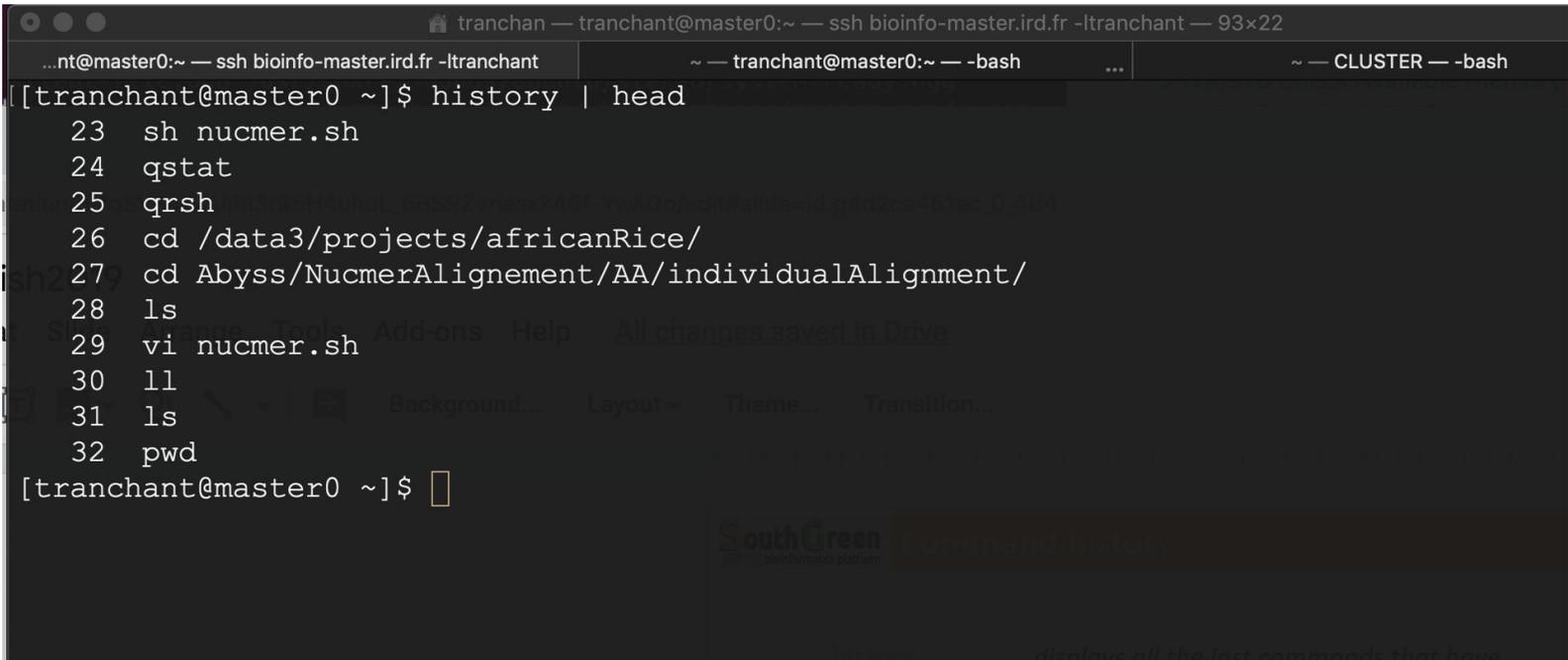


# Autres commandes utiles

## history

*afficher toutes les commandes enregistrées de la plus ancienne à la plus récente*

**L'historique est enregistré dans le fichier `.bash_history`**



```
tranchan — tranchant@master0:~ — ssh bioinfo-master.ird.fr -ltranchant — 93x22
...nt@master0:~ — ssh bioinfo-master.ird.fr -ltranchant  ~ — tranchant@master0:~ — -bash  ...  ~ — CLUSTER — -bash
[tranchant@master0 ~]$ history | head
 23  sh nucmer.sh
 24  qstat
 25  qrsh
 26  cd /data3/projects/africanRice/
 27  cd Abyss/NucmerAlignement/AA/individualAlignment/
 28  ls
 29  vi nucmer.sh
 30  ll
 31  ls
 32  pwd
[tranchant@master0 ~]$
```

## history

*afficher toutes les commandes enregistrées de la plus ancienne à la plus récente*

**L'historique est enregistré dans le fichier `.bash_history`**

## Filtering the History Output

```
history | head -n 5
```

Liste les commandes les plus anciennes

```
history | tail
```

Liste les commandes les plus récentes

```
history | grep "blastn"
```

Recherche les anciennes commandes "blastn"

```
history | grep "blastn" | tail -n 5
```

## Commande ls -l

```
$ ls -l filename  
drwxrwxrwx 3 user user 4096 2012-02-11 20:21 file_name
```

Permissions

Proprio

Groupe

Taille

Heure et date de la dernière modification

Type

-Interprétation/Légendes des permissions

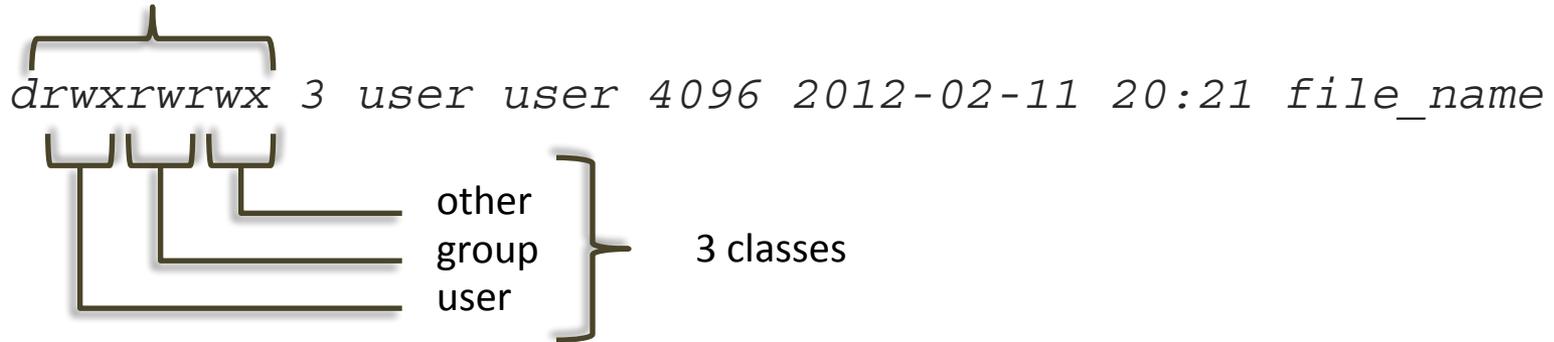
- : fichier normal

**d** : répertoire

**l** : lien symbolique

## ls -l command

Permissions



### 3 types of permissions :

Permission	File	Directory
Read r	Ouverture et Lecture	Lister et copier les fichiers
Write w	Modiier et supprimer	Manipuler le contenu : copier, créer, modifier, écraser
Execution x	Executer le fichier	Accès seulement au fichier pour l'exécuter

commande pour la gestion des permissions : **chmod**

```
chmod <perm> file_name
```

Chaque permission = 1 valeur

<b>R</b>	<b>4</b>
<b>W</b>	<b>2</b>
<b>X</b>	<b>1</b>
none	0

## Exemple

```
chmod 740 script.sh
```

```
# Owner=rwx Group=r-- Other=---
```

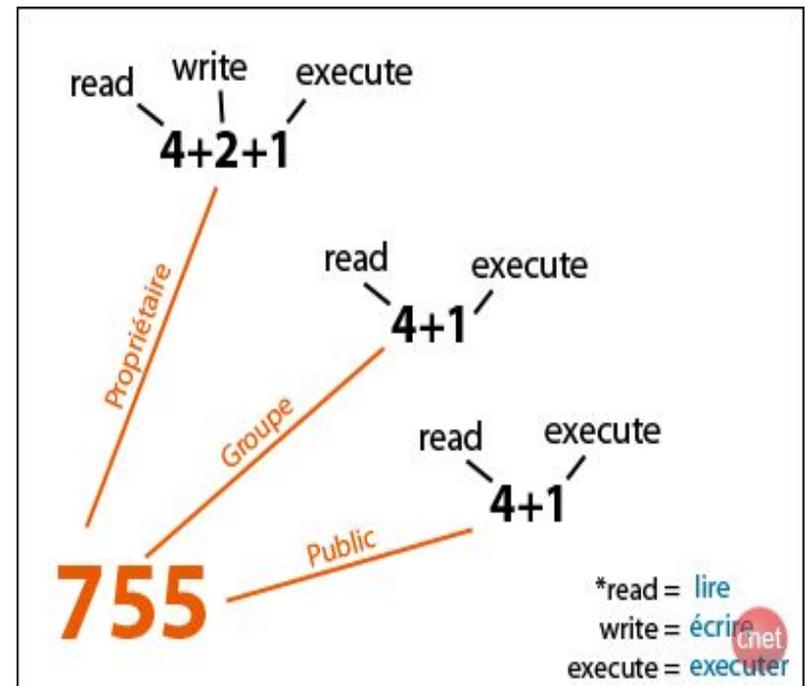
```
chmod 755 script.sh
```

```
# Owner=rwx Group=r-x Other=r-x
```

## chmod, ls

Donner le nom du propriétaire, du groupe et les droits des fichiers contenus dans le répertoire “~/Data/454-projet1/raw”

Changer les droits du fichier Scripts/blast.pl pour qu’il soit :  
en lecture/écriture pour le groupe,  
lecture/écriture/exécution  
pour le propriétaire  
et lecture au public



## Compresser des fichiers **tar,gzip**

```
tar -zcvf tarfile.tar.gz dirToCompress  
gzip fileToCompress
```

## Décompresser une archive **gunzip, tar**

```
gunzip file.gzip  
tar -xvf file.tar  
tar -zxvf file.tar.gz  
gzip -d file.gz
```

## Afficher le contenu d'une archive **zcat**

```
zcat data.txt.gz
```

## Rechercher une expression/motif dans une archive **zgrep**

```
zgrep 'NM_000020' data.gz
```

## rename

Exemple	Description
<code>rename 's/.txt/.fasta/' *</code>	rename the extension of all files
<code>rename 'y/a-z/A-Z/' *</code>	rename files in uppercase

Taille du disque **df**

disk free

```
df
```

taille espace occupé en octet

```
df -h
```

human-readable

Taille des répertoires **du**

disk usage

```
du
```

```
du -h
```

```
du -h *
```

Rechercher un fichier **find**

```
find -name "transcritsAssembly.fasta"
```

- **Christine Tranchant-Dubreuil**



- Sebastien Ravel



- Alexis Dereeper



- **Ndomassi Tando**



- François Sabot



- Gautier Sarah



- Bruno Granouillac



- **Valérie Noël**



- **Guilhem Sempere**



- **Florentin Constancias**



# Merci pour votre attention !



Le matériel pédagogique utilisé pour ces enseignements est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions (BY-NC-SA) 4.0 International:

<http://creativecommons.org/licenses/by-nc-sa/4.0/>