

Langages de scripts

Frédéric Goualard

Laboratoire des Sciences du Numérique de Nantes, UMR CNRS 6004
Office #112-11

FASTA

```
>P01013 GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDITLVLVNAIFYRGMWKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNVATLPAEKMKILELPFASGDLMLVL
PDEVSDLERIEKTIKFEKTEWNTWNTMEKRRVKVYLPQMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSED
GIEMAGSTGVIEDIKHSPSEQFRADHPFLFLIKHNPTNTITVYFGRYWSP
```



GenBank flatfile report

```
1      qikdllvsss  tdldttlvlv  naiyfkgmwk  tafnaedtre  mpfhvtkqes  kpvqmmcmnn
61     sfnvatlpae  kmkilelpfa  sgdlsmvlvl  pdevsdleri  ektinfeklt  ewtnpntmek
121    rrkvylpqm   kieekynlts  vimalgmtdl  fipsanltgi  ssaeslkisq  avhgafmels
181    edgiemagst   gviedikhsp  eseqfradhp  flflikhnpt  ntivyfgrywsp
```

FASTA

```
>P01013 GENE X PROTEIN (OVALBUMIN-RELATED)
QIKOLLVSSSTDLDITLVLVNAIFYFGMKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNVAATPAEKMKILELFPASGDLMLVLL
PDEVSDLERIEKTIKFEKTEWNNWMEKRRVKVYLPQMKIEEKYNLTSLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSED
GIEMAGSTGVIEDIKHSPSEQFRADHPFLFLIKHNPTNTIVYFGRYWSP
```



GenBank flatfile report

```
1      qikdllvsss  tdldttlvlv  naiyfgmwwk  tafnaedtre  mpfhvtkqes  kpvqmmcmnn
61     sfnavatpae  kmkilelpfa  sgdlsmlvll  pdevsdleri  ektinfeklt  ewtnpntmek
121    rrvkvylpqm  kieekynlts  vimalgmtld  fipsanltgi  ssaeslkisq  avhgafmels
181    edgiemagst  gviedikhsp  eseqfradhp  flflihnpt  ntivyfgrywsp
```

Solution BASH :

```
#!/usr/bin/env bash
tail -n +2 fasta.in | fold -w 60 |
tr A-Z a-z |
sed 's/\(.{10}\)/ \1/g' |
awk \
'{
printf("%3d\t%s\n",1+60*(NR-1), $0);
}'
```

Solution Python :

```
#!/usr/bin/env python3

from Bio import SeqIO
from Bio.Alphabet import generic_dna
with open("fasta.in", "rU") as f_in,\
     open("genbank.out", "w") as f_out:
    sequences = list(SeqIO.parse(f_in, "fasta"))
    for seq in sequences:
        seq.seq.alphabet = generic_dna
        SeqIO.write(sequences, f_out, "genbank")
```

Problème : déplacer chaque fichier GenBank **.gbk** d'un répertoire dans un répertoire avec le même nom (exemple : **AF068625.gbk** est déplacé dans **AF068625**).

Problème : déplacer chaque fichier GenBank **.gbk** d'un répertoire dans un répertoire avec le même nom (exemple : **AF068625.gbk** est déplacé dans **AF068625**).

Solution BASH :

```
for f in *.gbk; do
    mkdir "${basename "$f"}.gbk" && mv "$f" "$_"
done
```

Solution Python :

```
#!/usr/bin/env python3

import os
import shutil
from glob import glob

for fic in glob("*.gbk"):
    dname = os.path.splitext(fic)[0]
    os.mkdir(dname)
    shutil.move(fic,dname+"/"+fic)
```

Problème : déplacer chaque fichier GenBank **.gbk** d'un répertoire dans un répertoire avec le même nom (exemple : **AF068625.gbk** est déplacé dans **AF068625**).

Solution BASH :

```
for f in *.gbk; do
    mkdir $(basename "$f" .gbk) && mv "$f" "$_"
done
```

Solution Python :

```
#!/usr/bin/env python3

import os
import shutil
from glob import glob

for fic in glob("*.gbk"):
    dname = os.path.splitext(fic)[0]
    os.mkdir(dname)
    shutil.move(fic,dname+"/"+fic)
```

Objectif du cours : Savoir utiliser les langages de script au mieux de leurs capacités pour automatiser les tâches de collecte, gestion et analyse des données

- ▶ Introduction aux systèmes d'exploitation
- ▶ Langages de *shell* : BASH
 - ▶ + nombreuses applications (sed, grep, ...)
- ▶ Bases du langage Python

- ▶ Étude *personnelle* de chaque partie en dehors des créneaux de TD/TP
- ▶ Madoc : lire ou regarder les ressources fournies pour chaque partie
- ▶ Quiz sur madoc pour valider la compréhension de chaque partie.
- ▶ Créneaux de TD (5 créneaux)
 - ▶ Debriefing sur le quiz précédent et sur les notions correspondantes
 - ▶ Évaluation sur table d'une partie vue précédemment
- ▶ Créneaux de TP (12 créneaux)
 - ▶ Debriefing éventuel sur le quiz précédent
 - ▶ Exercices d'application plus fouillés (à préparer le plus possible à l'avance)
 - ▶ Évaluation sur ordinateur d'une partie vue précédemment
- ▶ Tous les détails sur le cours : **syllabus sur madoc (à lire impérativement)**
- ▶ **Prise en main de votre formation — rôle actif**

X1BI030

2022/2023

v. 2, 2022/09/01 / 5 TDs, 12 TP	Quiz (à faire avant le dimanche de la semaine précédente)	Créneau 1	Créneau 2	Créneau 3
(36) 05/09 – 09/09		TD1 Présentation du cours Intro. aux S.E. Introduction à BASH	TP1 Exercices BASH	
(37) 12/09 – 16/09	Introduction aux S.E. BASH 1	TP2 Discussion quiz S.E.+BASH 1 Exercices BASH	TP3 Exercices BASH	
(38) 19/09 – 23/09	BASH 2	TD2 BASH Discussion quiz BASH 2	TP4 Exercices BASH	
(39) 26/09 – 30/09				
(40) 03/10 – 07/10		TD3 Introduction à Python	TP5 Exercices Python 1	TP6 Exercices Python 1
(41) 10/10 – 14/10				
(42) 17/10 – 21/10	Python 1 Python 2	TD4 Évaluation S.E. et BASH Discussion quiz Python 1 & 2 Exercices Python	TP7 Évaluation BASH Exercices Python 2	
(43) 24/10 – 28/10		TD5 Python 2	TP8 Exercices Python + BASH	
(44) 31/10 – 04/11	Distribution du sujet de projet			
(45) 07/11 – 11/11				
(46) 14/11 – 18/11		TP9 Projet	TP10 Projet	
(47) 21/11 – 25/11		TP11 Évaluation BASH + Python	TP12 Projet	

- ▶ Une évaluation sur table E_1 de 30 minutes sur 10 points ;
- ▶ Deux évaluations sur machine E_2 et E_3 de 40 minutes sur 10 points ;
- ▶ Un projet de TP sur 20 points : T

Note finale :

$$N = \frac{2}{5}(E_1 + E_2 + E_3 + T)$$