

# 1)INTRODUCTION

## 2)Présentation de l'organisme

### 2.1)présentation générale

- Université d'Auvergne (UdA): 16 000 étudiants, 3 000 étrangers

3 Écoles universitaires (Droit,Économie,Management); 3 Facultés (Médecine,Pharmacie,Chirurgie Dentaire)  
IUT (Biologie,Informatique (Informatique,réseaux et multimédia), SI (Sciences pour l'ingénieur),Gestion)  
22 labos dont EA CIDAM

- EA CIDAM: <http://pharmacie.u-clermont1.fr/ea-cidam.html>

### 2.2)conditions du stage

- locaux
- equipe

## 3)Moyens techniques

### 3.1)Outils bioinformatiques utilisés

#### 3.1.1)ASaiM

- presentation générale
- environment
- framework
- interface web

#### 3.1.2)Galaxy

- plate-forme open source
- accessible en ligne + téléchargeable en local
  - intérêt en ligne
    - outils installés → utilisation rapide
    - puissance → tourne avec données taille conséquentes
  - intérêt local
    - sélectionner propres outils
    - pas de file d'attente → outils et inputs de petite taille
- analyses simples(entrer params), complètes et reproductibles
- resultats=datasets

#### 3.1.3)Wrapper

- wrapper :
  - interface, packaging outil

- « appel » automatique de l'outil et des packages requis
- code xml :
  - <tool> + <description>
  - <requirements> + <version\_command>
  - <command>
  - <inputs>
  - <outputs>
  - <tests>
  - <help> + <citations>
- Le test dans les wrappers
  - part du principe que l'outil sur le quel on travail fonctionne correctement
  - test wrapper doit utiliser au moins une fois chaque paramètre
  - utilise fichiers de tests : de petite taille, même caract éristiques que fichier normal, données réalistes
  -

### **3.2)"Techniques utilisées"**

#### **3.2.1)"Travail en collaboration"**

- git
  - branches indépendantes + arborescence
  - commit
  - accès anciens commits
- github
  - PR
  - collaborateurs

#### **3.2.2)planemo**

- presentation planemo
- commandes planemo
  - tool\_init
  - lint + shed\_lint
  - test
  - shed\_create + shed update

#### **3.2.3)Travis CI**

- presentation travis
- intérêt
- travis.yml

## **4)Travaux réalisés**

### **4.1)outils**

- presentation certains outils
  - combine\_metaphlan2\_humann2
  - extract\_min\_max\_lines
  - fasta\_add\_barcode
  - extract\_sequence\_file
  - normalize\_dataset
  - format\_metaphlan2\_output
  - outils plot
    - plot\_generic\_x\_y\_plot
    - plot\_barplot
    - plot\_grouped\_barplot
  - cdhit
    - cd\_hit\_est
    - cd\_hit\_protein
    - format\_cd\_hit\_output
  - graphlan
    - export2graphlan
    - graphlan
    - graphlan\_annotate
  - humann2
    - humann2\_join\_tables
    - humann2\_reduce\_table
    - humann2\_regroup\_table
    - humann2\_rename\_table
    - humann2\_renorm\_table
    - humann2\_split\_table
    - compare\_humann2\_output

## **4.2)Pb rencontrés**

- fichiers de test manquants
- problèmes liés a galaxy
  - fichiers téléchargés renommés data.. ex sortie join\_tables
- spécificités liées aux outils : formats des fichiers
- paramètres optionnels : rename table
- version d'un package : ex graphlan/tool\_dependencies.xml
- test ne pouvant exister : humann2\_download

## **4.3)Solution**

- nouveaux fichiers inputs : ex split table
  - fichiers de test fourni par l'outil
- utilisation des outils pour obtenir les sorties
- modification des wrappers
  - ex pour param optionnels : integration boucle if
  - etc

## **5) Bilan**

### **5.1) Bilan professionnel**

### **5.2) Bilan personnel**