

DAZENIERE Julie  
KIM Yujin  
MORANGES Maëlle



# **Projet *Dickeya* : Guide d'utilisateur**



M1 Bioinformatique moléculaire : méthodes et analyses

# Table des matières

|  |   |
|--|---|
| 1) Page d'accueil .....  | 3 |
| 2) Page Requêtes .....   | 3 |
| a. Bloc 1 : sélection des souches d'intérêt .....  | 4 |
| b. Bloc 2 : sélection d'une fonction d'intérêt .....   | 5 |
| c. Bloc 3 : sélection d'une fonction d'intérêt .....   | 5 |
| d. Bloc 1 et 2 : sélection des souches d'intérêt présentant une<br>fonction sélectionnée .....                             | 5 |
| e. Bloc 1 et 3 : sélection des souches d'intérêt présentant certaines<br>contraintes .....                                 | 5 |
| f. Bloc 2 et 3 : sélection d'une fonction et de contraintes .....  | 6 |
| g. Bloc 1, 2 et 3 : sélection des souches d'intérêt présentant une<br>fonction spécifique, ainsi que des contraintes ..... | 6 |
| 3) Page de résultat .....  | 6 |

## 1) Page d'accueil

Cette page permet à l'utilisateur de sélectionner le projet sur lequel il souhaite travailler. En l'état, seul le projet « Dickeya » peut être étudié.



**INSA** INSTITUT NATIONAL  
DES SCIENCES  
APPLIQUÉES  
LYON

**Dickeya** Lyon 1

🏠

**Accueil**

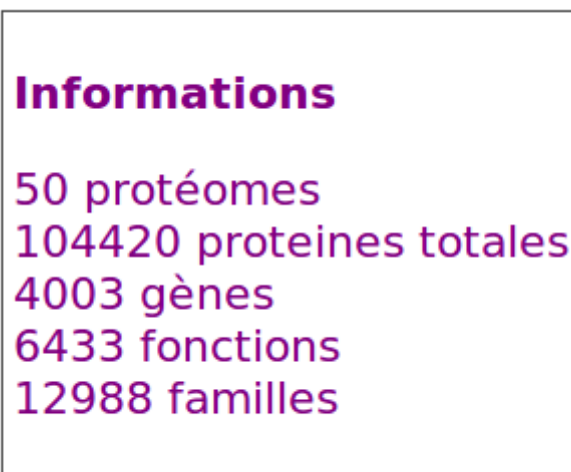
Sélectionner un projet :

Dickeya ▼

Visualiser ce projet

## 2) Page Requêtes

A droite vous trouverez un résumé des données présentes dans ce projet.



**Informations**

50 protéomes  
104420 protéines totales  
4003 gènes  
6433 fonctions  
12988 familles

Sur la page de requêtes se trouvent 3 blocs permettant d'effectuer des requêtes. Il est possible d'exécuter chaque requête une à une (a, b, c), mais également de les combiner (d, e, f, g). L'utilisateur doit cocher les blocs correspondant à la requête qu'il souhaite effectuer.

### a. Bloc 1 : sélection des souches d'intérêt

Ce premier bloc permet de visualiser la liste des protéines contenues dans une ou plusieurs souches, en cochant les souches d'intérêt dans la première zone de cases à cocher.

Il est également possible de choisir de visualiser les protéines présentes dans un set de protéomes, mais pas dans un autre.

Description des boutons :

- « Tout cocher » (1) : permet de cocher toutes les souches qui ne sont pas cochées dans « absentes dans » (zone rouge).
- « Tout décocher » (2) : permet de décocher toutes les souches cochées dans la zone 1.
- « Cocher tous les autres » (3) : permet de cocher toutes les souches qui ne sont pas cochées dans « présentes dans » (zone bleue).
- « Tout décocher » (4) : permet de décocher toutes les souches cochées dans la zone 2.

## Requêtes

### ☒ Protéines présentes dans ces protéomes :

- ☐ Cedecea\_neteri\_ND14b
- ☒ Dickeya\_aquatica
- ☒ Dickeya\_aquatica\_DW\_0440
- ☒ Dickeya\_chrysanthemi\_Ech1591
- ☐ Dickeya\_chrysanthemi\_L11
- ☒ Dickeya\_chrysanthemi\_M074
- ☐ Dickeya\_chrysanthemi\_NCPRB\_3533

1

Tout cocher

2

Tout décocher

### Mais absentes dans ceux-ci:

- ☒ Cedecea\_neteri\_ND14b
- ☐ Dickeya\_aquatica
- ☐ Dickeya\_aquatica\_DW\_0440
- ☐ Dickeya\_chrysanthemi\_Ech1591
- ☒ Dickeya\_chrysanthemi\_L11
- ☐ Dickeya\_chrysanthemi\_M074
- ☒ Dickeya\_chrysanthemi\_NCPRB\_3533

3

Cocher tous les autres

4

Tout décocher

### *b. Bloc 2 : sélection d'une fonction d'intérêt*

Le deuxième bloc permet de visualiser la liste des protéines impliquées dans une fonction sélectionnée grâce à la liste déroulante.

☒ **Protéines impliquées dans la fonction**

1,3-propanediol dehydrogenase

(+)-neomenthol dehydrogenase  
(2,3-dihydroxybenzoyl)adenylate synthase  
(2E,6E)-farnesyl diphosphate synthase  
(2E,6E)-farnesyl- diphosphate-specific ditrans,polycis-undecaprenyl-diphosphate synthase  
(2Fe-2S) ferredoxin  
(2Fe-2S)-binding protein  
(4Fe-4S)-binding protein  
(alpha)-aspartyl dipeptidase  
(Fe-S)-cluster assembly protein  
**(p)ppGpp synthetase**  
1,3-beta-glucan synthase regulator  
1,3-propanediol dehydrogenase  
1,4-alpha-glucan branching enzyme  
1,4-beta-xylanase  
1,4-dihydroxy-2-naphthoate octaprenyltransferase  
1,4-dihydroxy-2-naphthoate polyprenyltransferase  
1,4-dihydroxy-2-naphthoyl-CoA synthase  
1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase  
1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase  
1-acyl-sn-glycerol-3-phosphate acyltransferase

### *c. Bloc 3 : sélection d'une fonction d'intérêt*


Le troisième bloc permet de filtrer le résultat des requêtes selon 3 paramètres :

- Le pourcentage d'identité supérieur à un certain seuil (par défaut 0%)
- Le pourcentage de gap inférieur à un certain seuil (par défaut 100%)
- La taille de l'alignement en acides aminés, supérieure à un certain seuil (par défaut 0)

☒ **Paires de protéines telles que:**

Pourcentage identité supérieur à :  %

Pourcentage de gap inférieur à :  %

Taille alignement plus grand que :   acides aminés

### *d. Bloc 1 et 2 : sélection des souches d'intérêt présentant une fonction sélectionnée*

Cette combinaison permet ainsi de sélectionner les protéomes possédant des protéines impliquées dans une fonction spécifique. Comme pour la requête 1 simple, il est possible d'exclure des souches.

### *e. Bloc 1 et 3 : sélection des souches d'intérêt présentant certaines contraintes*

Cette combinaison permet de retourner les paires protéines présentes dans un set de souches, et correspondant aux paramètres renseignés.

### *f. Bloc 2 et 3 : sélection d'une fonction et de contraintes*

Cette combinaison permet de renvoyer les paires de protéines possédant la fonction choisie, ainsi que les paramètres renseignés.

### *g. Bloc 1, 2 et 3 : sélection des souches d'intérêt présentant une fonction spécifique, ainsi que des contraintes*

Cette requête permet de sélectionner les paires de protéines présentes dans un set de souches (possibilité d'exclure un autre set), possédant une fonction choisie, et respectant les contraintes indiquées.

## 3) Page de résultat

Sur cette page vous trouverez un rappel de la requête envoyée dans la page précédente.

Ensuite sont affichés les 1000 premiers résultats de la requête. Ces derniers sont ordonnés dans un ordre spécifique selon la requête :

- Alphabétique de protéines et de souches pour les requêtes 1 et 2.
- Si la requête est impliquée, l'ordre est le suivant :
  - o Pour les protéines identiques : nom de la protéine, puis de la souche 1, et enfin la souche 2.
  - o Pour les protéines similaires : pourcentage d'identité, pourcentage de gap, nom de la souche 1, nom de la protéine 1, puis nom de la souche 2 et de la protéine 2.

Si la requête 3 n'est pas cochée, les résultats s'affichent dans un seul tableau :

**Résultats**

Télécharger résultat dans :  .txt

**Rappel de la requête :**  
Protéines présentes dans ('Dickeya\_aquatica\_DW\_0440','Dickeya\_solani\_IPO\_2222') mais pas dans (') et possédant la fonction 30S ribosomal protein S18

| Noms                     | NomP                | NomGene | NomFo                     |
|--------------------------|---------------------|---------|---------------------------|
| Dickeya_aquatica_DW_0440 | WP_000135199.1 rpsR |         | 30S ribosomal protein S18 |
| Dickeya_solani_IPO_2222  | WP_000135199.1 rpsR |         | 30S ribosomal protein S18 |

Julie Dazenièrre, Yujin Kim & Maëlle Moranges [M1 Bio-info](#) 2017

Si elle est cochée, les résultats sont dans deux tables :

- La première correspond aux protéines identiques, c'est-à-dire qu'une ligne correspond à une même protéine partagée par deux souches différentes. Du coup les pourcentages d'identité, de gap, et la taille de l'alignement ne sont pas affichés car ils possèdent toujours les valeurs suivantes : 100% d'identité, 0% de gap et taille de l'alignement correspondant à la longueur de la protéine.
- La deuxième table correspond aux protéines similaires : il s'agit donc de protéines différentes, respectant les conditions indiquées dans la requête.

**Rappel de la requête :** Protéines présentes dans ('Dickeya\_chrysanthemi\_Ech1591','Dickeya\_chrysanthemi\_NCPPB\_3533') mais pas dans ('') et présentant les contraintes suivantes :

- %gap < 100
- %id > 75
- taille de l'alignement > 0 acides aminés

**Protéines identiques :**

| NomSouche1                      | NomProteine1   | NomSouche2                   | NomProteine2   | Gene | Fonction                           |
|---------------------------------|----------------|------------------------------|----------------|------|------------------------------------|
| Dickeya_chrysanthemi_NCPPB_3533 | WP_000135199.1 | Dickeya_chrysanthemi_Ech1591 | WP_000135199.1 | rpsR | 30S ribosomal protein S18          |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_001144069.1 | Dickeya_chrysanthemi_Ech1591 | WP_001144069.1 | rpsU | 30S ribosomal protein S21          |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_001157747.1 | Dickeya_chrysanthemi_Ech1591 | WP_001157747.1 |      | DNA-binding response regulator     |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_001181005.1 | Dickeya_chrysanthemi_Ech1591 | WP_001181005.1 | rpsJ | 30S ribosomal protein S10          |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_002208627.1 | Dickeya_chrysanthemi_Ech1591 | WP_002208627.1 | glnK | nitrogen regulatory protein P-II 2 |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_002211347.1 | Dickeya_chrysanthemi_Ech1591 | WP_002211347.1 | infA | translation initiation factor IF-1 |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_002919257.1 | Dickeya_chrysanthemi_Ech1591 | WP_002919257.1 | rpsK | 30S ribosomal protein S11          |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_004093904.1 | Dickeya_chrysanthemi_Ech1591 | WP_004093904.1 | atpE | ATP synthase subunit C             |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_004846568.1 | Dickeya_chrysanthemi_Ech1591 | WP_004846568.1 | rpmD | 50S ribosomal protein L30          |
| Dickeya_chrysanthemi_NCPPB_3533 | WP_005972168.1 | Dickeya_chrysanthemi_Ech1591 | WP_005972168.1 | rsmA | carbon storage regulator           |

**Protéines similaires :**

| Souche1                      | Proteine1      | Gene1 | Fonction1                 | Souche2                         | Proteine2      | Gene2 | Fonction2                 | PourcentageId | PourcentageGap | TailleAli |
|------------------------------|----------------|-------|---------------------------|---------------------------------|----------------|-------|---------------------------|---------------|----------------|-----------|
| Dickeya_chrysanthemi_Ech1591 | WP_009113827.1 | cspE  | cold-shock protein        | Dickeya_chrysanthemi_Ech1591    | WP_012768004.1 |       | cold-shock protein        | 75            | 0              | 68        |
| Dickeya_chrysanthemi_Ech1591 | WP_009113827.1 | cspE  | cold-shock protein        | Dickeya_chrysanthemi_NCPPB_3533 | WP_012768004.1 |       | cold-shock protein        | 75            | 0              | 68        |
| Dickeya_chrysanthemi_Ech1591 | WP_012768184.1 |       | transcriptional regulator | Dickeya_chrysanthemi_Ech1591    | WP_015848023.1 |       | transcriptional regulator | 75            | 0              | 76        |
| Dickeya_chrysanthemi_Ech1591 | WP_012768184.1 |       | transcriptional           | Dickeya_chrysanthemi_NCPPB_3533 | WP_015848023.1 |       | transcriptional           | 75            | 0              | 76        |

Il est également possible d'enregistrer la totalité des résultats dans un fichier tabulé au format .txt. Ce fichier peut ensuite être ouvert avec un tableur (ex : Excel, Classeur de LibreOffice), de manière à obtenir un tableau. L'utilisateur doit renseigner le nom du fichier, appuyer sur « ok ». Un lien vers le ou les fichiers générés est alors affiché. Lorsque l'utilisateur clique sur le lien, une nouvelle page s'ouvre avec les résultats obtenus. Il est possible d'enregistrer ces résultats dans un fichier en faisant un click droit → enregistrer sous...

Si la requête 3 n'est pas cochée, tous les résultats sont présents dans un seul fichier.

## Résultats

Télécharger résultat dans :  .txt  [example2.txt](#)

Dans le cas contraire, vous trouverez deux fichiers : l'un contenant les protéines identiques, et l'autre les protéines similaires.

## Résultats

Télécharger résultat dans :  .txt  [example\\_similaires.txt](#) [example\\_identiques.txt](#)

## **/!\ Attention :**

La requête 3 est très longue à effectuer, c'est pourquoi nous affichons peu de résultats (10 protéines identiques et 10 protéines similaires) sur la page résultat. L'enregistrement des résultats de la requête 3 seule ou une combinaison de la requête 1 et 3 va être très longue, dans l'ordre des heures avec les paramètres les moins stringents possibles (un grand nombre de protéomes par exemple).