



DESS **Compétence Complémentaire en Informatique**
Septembre 2002

Soutenance de stage



Jean MULLER



**Développement d'un logiciel
bioinformatique d'analyse de groupes de
protéomes et de génomes :**

**Application aux bactéries pathogènes
impliquées dans les diarrhées humaines**

Soutenance de stage

Titre

Jean Muller



Déroulement général

- La pathologie
- Le projet réalisé durant le stage
- Un exemple de résultat

Soutenance de stage

Jean Muller



La pathologie

- La diarrhée compte parmi les causes de mortalité les plus importantes chez les enfants de moins de 5 ans
- D'origine bactérienne, elle est liée au pouvoir pathogène des bactéries, sous la forme de toxines ou d'autres protéines
- La détermination de ses causes constitue un enjeu très important

Soutenance de stage

La pathologie

Jean Muller

Approche génomique

- Les génomes de plus de 70 organismes différents ont été complètement séquencés (~50 bactéries)
- ADN : le génome (~4 millions de bases pour bactérie)
protéines : le protéome (~4000 protéines)
- Biologie et traitements informatiques à haut débit (automatisation, banques de données)
- Études comparatives des protéomes des bactéries pathogènes

Soutenance de stage

Génomique

Jean Muller

Sur quels génomes ?

DiaBac
Available complete genomes of BACteria that cause DIArthra

- Bactéries responsables de la diarrhée chez l'homme
- 9 bactéries dont le génome est complètement séquencé

Organisme	Souche	Abréviation
<i>Campylobacter jejuni</i>	NCTC 1116	CJEJ
<i>Escherichia coli</i>	K12- M61655	ECOL
<i>Escherichia coli</i>	O157:H7	ECOH
<i>Helicobacter pylori</i>	26695	HPYL
<i>Helicobacter pylori</i>	J99	HPYJ
<i>Staphylococcus aureus</i>	N315	SAUR
<i>Salmonella enterica</i>	CT18	SENT
<i>Salmonella typhimurium</i>	NCTC 1116	STYM
<i>Vibrio cholerae</i>	Biotype El Tor	VCHO

Soutenance de stage

Le stage

Jean Muller

Génomes et Diarrhée

- Caractériser les protéines responsables !
- Protéines spécifiques aux bactéries pathogènes ?
- Profil de présence/absence des protéines homologues aux protéines de DiaBac dans les autres organismes

Soutenance de stage

Le stage

Jean Muller

Homologie

- Deux protéines sont dites homologues lorsque au moins 20% de leurs séquences sont identiques
- Ceci est généralement lié à leur évolution à partir d'un ancêtre commun
- L'homologie révèle souvent des fonctions proches
- Outil de prédilection de recherche dans les banques de données (logiciels Blast)

Soutenance de stage

Le stage

Jean Muller



Les étapes de l'étude

- Création et mise à jour des bases de données GScope
 - Création de la base
 - Vérification de la base
- Validation des protéomes
 - Détection de protéines non prédites ou mal prédites
- Profil de Présence/Absence des homologues des protéines DiaBac
 - Création de famille, Bilan X-HDA, Clustering

Soutenance de stage

Le stage

Jean Muller



GScope

- Outil d'analyse et de visualisation de génomes et de protéomes
 - Détermination des séquences
 - Recherche d'homologues dans les banques nucléiques et protéiques
 - Alignements multiples
 - Génomique comparative
- 40000 lignes de Tcl/Tk, 2000 procédures

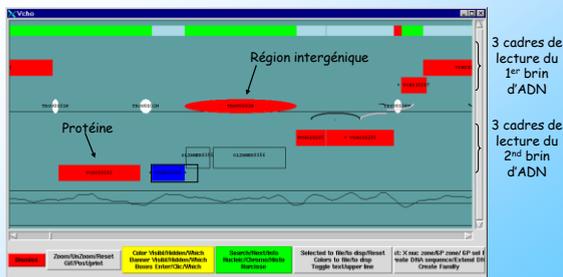
Soutenance de stage

Le stage

Jean Muller



Le Board de Gscope



3 cadres de lecture du 1^{er} brin d'ADN

3 cadres de lecture du 2nd brin d'ADN

Soutenance de stage

Le stage

Jean Muller



Base de données GScope

Une base Gscope pour un génome c'est, en automatique :

- Plus de 10000 fichiers créés
- Plus de 500 Mo de données
- 3 jours de calculs

Soutenance de stage

Le stage

Jean Muller



Les 9 génomes

- ➡ Création ou réactualisation des bases de données Gscope plus de 30000 protéines
- ➡ En automatique, mais ...
- ➡ « Inventory Check »
Test des étapes de la construction de la base Gscope

Soutenance de stage Le stage Jean Muller



Inventory Check

Pour exécuter les procédures illuminées : Control+Mouse1

Mise en évidence des étapes manquantes ou incomplètes, lors de la mise en place d'une base de données Gscope

Soutenance de stage Le stage Jean Muller



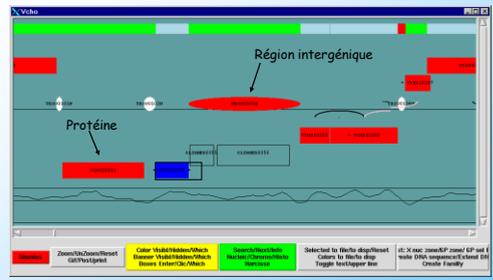
Validation des protéomes DiaBac

- ➡ Les protéomes sont prédits par des programmes non infallibles
- ➡ Protéines "ratées" ou limites incorrectes
- ➡ Augmenter la qualité des protéomes et disposer d'un matériel initial « propre »

Soutenance de stage Le stage Jean Muller



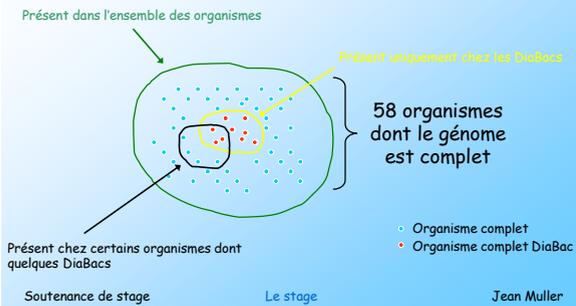
Validation



On a ainsi du, pour HpyJ par exemple, créer 4 protéines et en modifier 64

Soutenance de stage Le stage Jean Muller

Profil de Présence/Absence



HDA

Homologue Detection Agreement

Pour une **seule** protéine appartenant à un organisme

- Recherche dans les autres protéomes BlastP
- Recherche dans les autres génomes tBlastN
- Détection « plus ou moins valide » des homologues de cette protéine

Soutenance de stage

Le stage

Jean Muller

X-HDA

Crossed-linked Homologue Detection Agreement

L'intégration des 9 génomes de DiaBac permet d'établir des familles de protéines

- Recherches pour chaque protéine de la famille dans les protéomes et génomes (BlastP et tBlastN)
- Amélioration de la détection des familles de protéines homologues

Soutenance de stage

Le stage

Jean Muller

Création des Familles

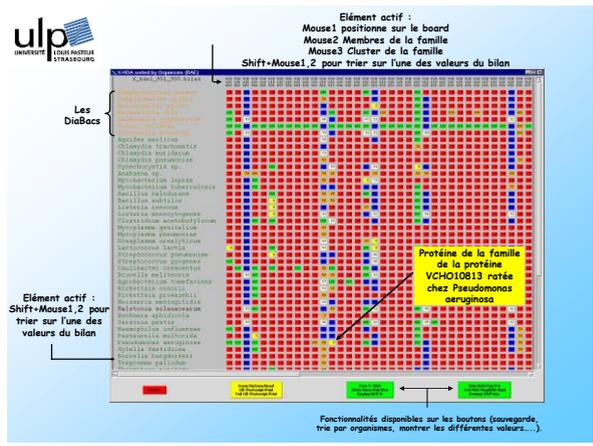
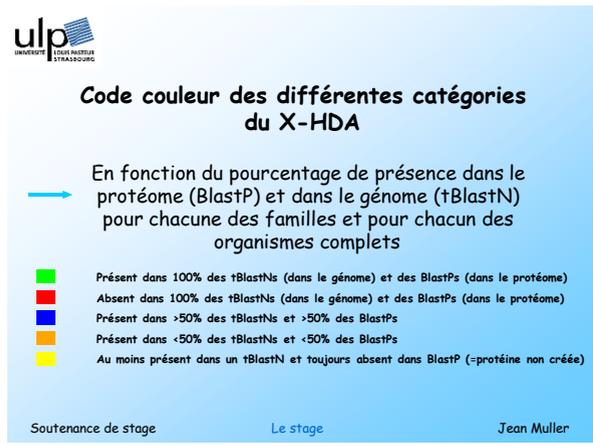
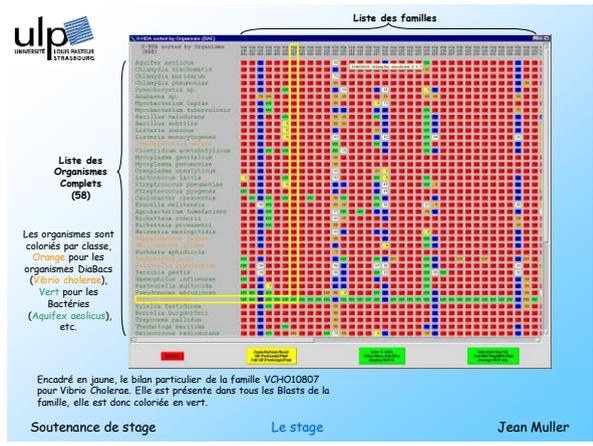
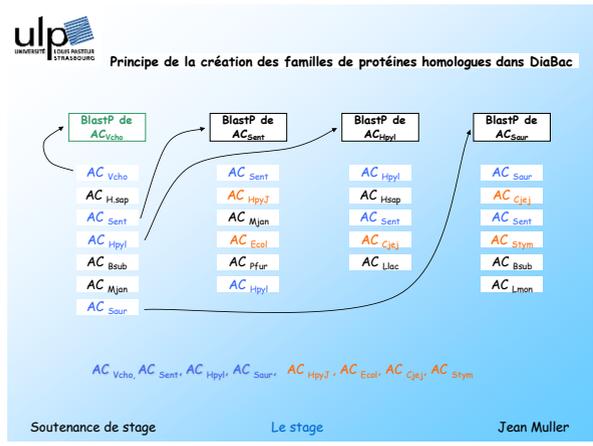
A partir d'une protéine de référence

- Recherche des homologues dans DiaBac
- Recherche des homologues des homologues dans DiaBac

Soutenance de stage

Le stage

Jean Muller





Alignement multiple

```

VCHO2002 - AminoAcidProperties/VCHO2002
VCHO2002 - AminoAcidProperties/VCHO2002

BAB37867  FENLSELEI  NQDFFTEIL  QGEGALQNP  QGCHALQVY  ...  SAALFA  DQVFIQDQ  LELQWQHP  ACTVDFQD  PFFICALL  TLVWLELD
AAAS0510  FENLSELEI  NQDFFTEIL  QGEGALQNP  QGCHALQVY  ...  SAALFA  DQVFIQDQ  LELQWQHP  ACTVDFQD  PFFICALL  TLVWLELD
P3768     FENLSELEI  NQDFFTEIL  QGEGALQNP  QGCHALQVY  ...  SAALFA  DQVFIQDQ  LELQWQHP  ACTVDFQD  PFFICALL  TLVWLELD
BAU21517  VEMLSQLEI  SQDFFTEIL  QGEGALQNP  QGCHALQVY  ...  CGLKFI  IAVVFIQDQ  IARELQHP  ADVDFQDQ  PFFICALLE  TVYVRELEA
VCHO2002  VTONALFIT  SQDFFTEIL  VQVFIQDQ  VQVFIQDQ  ...  LKAGK  VAQVFIQDQ  LQVFIQD  F  VQVFIQDQ  L  VAQVFIQD  LQVFIQDQ
QW042    VTONALFIT  SQDFFTEIL  VQVFIQDQ  VQVFIQDQ  ...  LKAGK  VAQVFIQDQ  LQVFIQDQ  F  VQVFIQDQ  L  VAQVFIQD  LQVFIQDQ
QW044    TAPVQKLEQ  SQDFFTEIL  SQDFFTEIL  SQDFFTEIL  ...  LKAGK  VAQVFIQDQ  LQVFIQDQ  F  VQVFIQDQ  L  VAQVFIQD  LQVFIQDQ
QW045    TFFVWVET  SQDFFTEIL  SQDFFTEIL  SQDFFTEIL  ...  LKAGK  VAQVFIQDQ  LQVFIQDQ  F  VQVFIQDQ  L  VAQVFIQD  LQVFIQDQ
QW046    PWS  FVLSQ  ELQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ
BAB37867  LQI  IFFIQDQ  LAAL  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ
AAAS0510  LQI  IFFIQDQ  LAAL  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ
P3768     LQI  IFFIQDQ  LAAL  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ
BAU21517  LQI  IFFIQDQ  LAAL  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ
VCHO2002  ELQVFIQDQ  LQI  IFFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ
QW042    ELQVFIQDQ  LQI  IFFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ
QW044    PFFVFIQDQ  KANAVFIQDQ  QWVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ
QW045    PFFVFIQDQ  KANAVFIQDQ  QWVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ  ...  LQVFIQDQ  LQVFIQDQ

```

Soutenance de stage Exemple Jean Muller



Exemples

- Cluster intéressant
Présence de protéines d'îlots pathogènes
 - Analyse VCHO20023 (hypothetical)
Homologue aux Racémas (modifications de la membrane bactérienne)
L'étude de l'alignement multiple montre qu'elle n'est constituée que de la partie transmembranaire
Il existe 2 formes de racémas : en 2 parties (interaction) et en 1 partie (fusionnelle). Les bactéries pathogènes semblent posséder les 2 alors que les autres seulement la fusionnelle.
 - La Racémas est-elle impliquée dans la pathogénicité ?
- Soutenance de stage Exemple Jean Muller



Perspectives

- Dans le cadre de la collaboration avec l'équipe de génomique structurale basée en Inde, caractériser d'autres protéines cibles pour la détermination de leur structure 3D
-
- ```

LCLVCSDEASGCHYVLTGSSCKVFFKRAVEQG
HNYLCAGRNDICIKRRKNCPCRYRKLQAG
MNLÉAR

```
- Parfaire l'étude détaillée des bilans X-HDA et des clusters disponibles
  - Poursuivre le développement logiciel
- Soutenance de stage      Jean Muller



## Bilan de stage

- Découverte des problématiques et des outils bioinformatiques
  - Programmation et intégration dans une plateforme existante
  - Application pratiques de résultats biologiques
- 50% informatique      50% biologie
- Soutenance de stage      Jean Muller



## Remerciements

- Enseignants du DESS CCI
- Dino Moras, Jean-Claude Thierry
- Olivier Poch
- Raymond Ripp et toute l'équipe

Soutenance de stage

Jean Muller