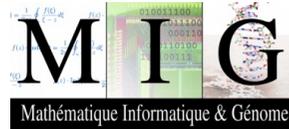


Web Services for Microbial Genome Annotation using Data Integration

C. MICHOTÉY *et al.*



CNRS UPMC INSU
Station Biologique
Roscoff



Unité
MIA Jouy en Josas
Mathématiques et Informatique Appliquées

Microbial Genome Annotation using Data Integration

Objectif :

Intégration de données pour l'aide à l'annotation de génomes microbiens

⇒ Architecture de type Web Services autour de 3 applications

Annotation de
génomes microbiens

The logo for AGMIAL, featuring the letters 'AGMIAL' in a serif font, with a small graphic of a DNA double helix behind the 'M'. The logo is contained within a yellow rectangular box with a folded bottom-right corner.The logo for MOSAIC, featuring the word 'MOSAIC' in a blue, bold, sans-serif font. The logo is contained within a purple rectangular box with a folded bottom-right corner.

Génomique
comparative

The logo for PARIS, featuring the word 'PARIS' in a green, bold, sans-serif font, with a small graphic of a DNA double helix behind the 'S'. The logo is contained within a green rectangular box with a folded bottom-right corner.

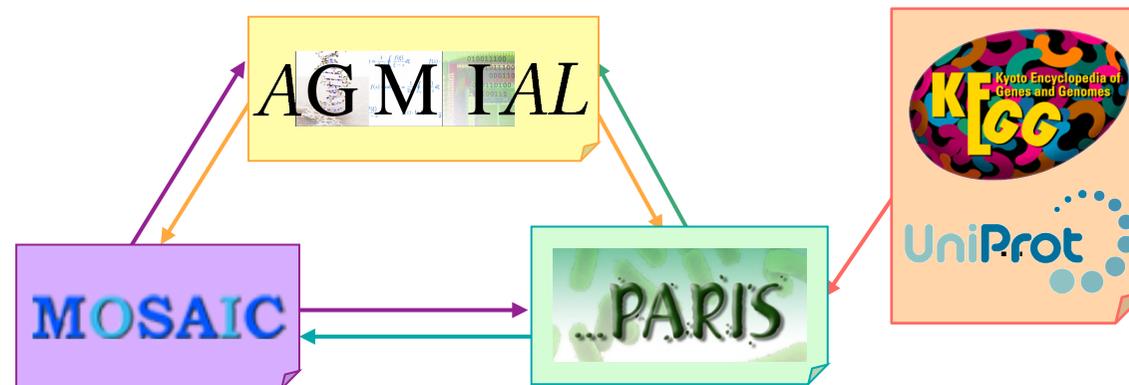
Protéomique
expérimentale

Microbial Genome Annotation using Data Integration

Objectif :

Intégration de données pour l'aide à l'annotation de génomes microbiens

⇒ Architecture de type Web Services autour de 3 applications

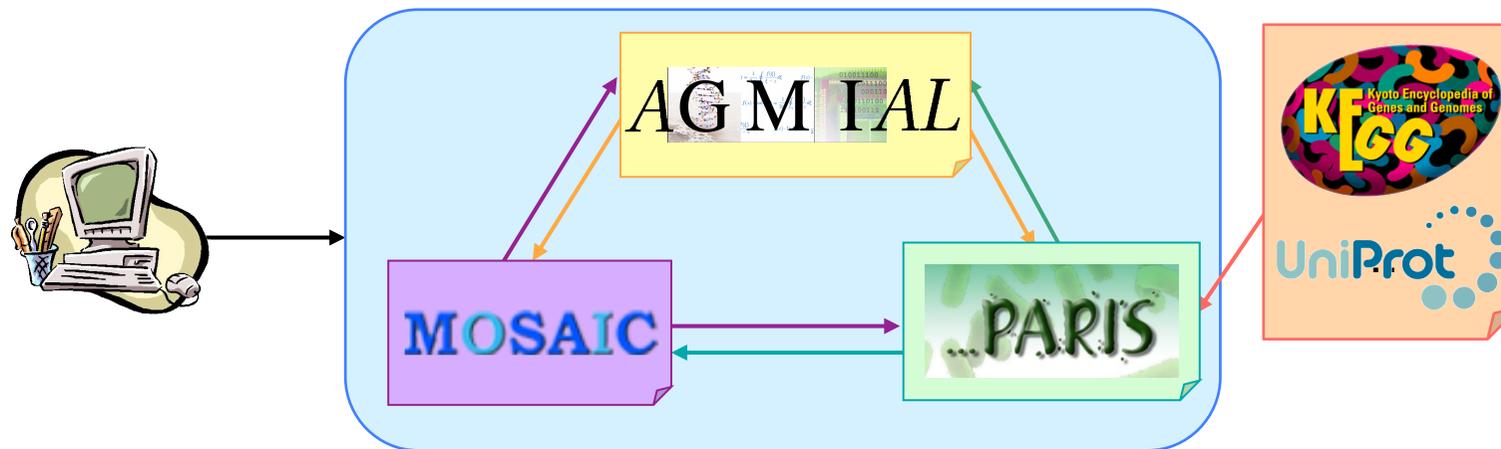


Microbial Genome Annotation using Data Integration

Objectif :

Intégration de données pour l'aide à l'annotation de génomes microbiens

⇒ Architecture de type Web Services autour de 3 applications



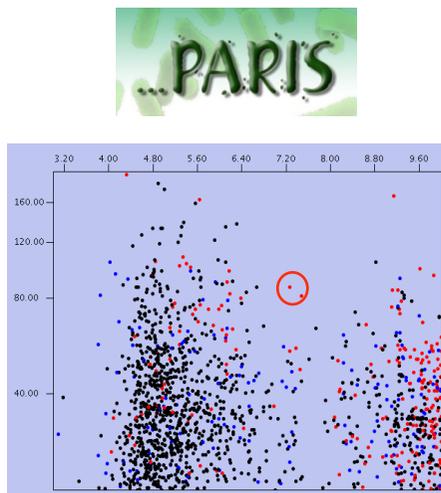
Méthodologie

- × Élaboration de scénarios en collaboration avec des microbiologistes
- × Définition d'une API (Application Programming Interface)
- × Développement de la couche Web Service
 - Chaque composant devient fournisseur de données
- × Extension des applications
 - Applications potentiellement clientes les unes des autres
- × Création d'un portail d'intégration
- × Données génomiques et protéomiques
 - *Staphylococcus aureus* (données publiques: Genbank et Expasy)
 - *Lactococcus lactis* (données publiques: INRA/PAPPSO)
 - *Propionibacterium freudenreichii* (données privées: INRA/STLO)

Exemple de scénario

Aide à l'interprétation des gels 2D

- × Protéine à expression variable dans un gel 2D
- × Position du gène correspondant sur le génome
 - Type de région (colinéaire/réarrangé/mixte)
 - Type du segment (Backbone Segment/Variable Segment/mixte)
- × Croisement avec régions régulatrices connues ?



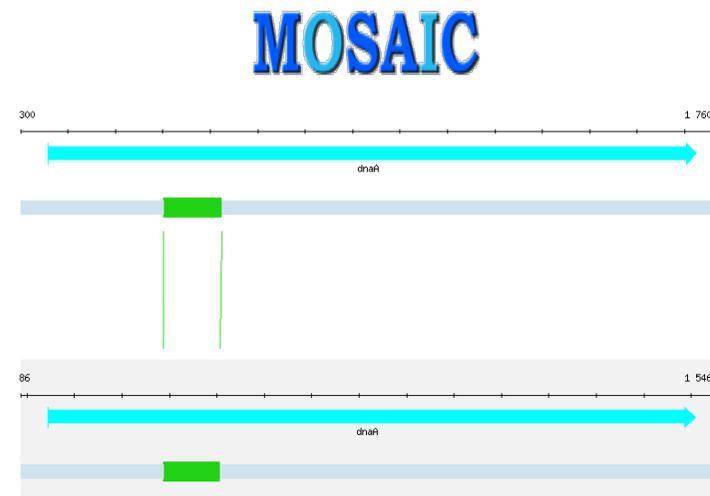
Gel 2D *L. lactis* IL1403



IL1403

L. lactis

SK11



Alignement *L. lactis* IL1403/SK11

Extension des interfaces

PARIS: Proteomic Analysis and Resources Indexation System [migadi connected on paris2_dev]

File Tools Preferences Configuration Help

Navigation Zoom no image. Zoom Control

Data

Spot	Protein	Gel	Agmial
SpotID			85
X			569
Y			275
MW			73000.0
PI			4.83
Relative volume			0.0
Identification			ms-maldi
Annotation mode			experiment
Area			0.0

Gene informations

Gene ID	Gene Name	Locus Tag
	dnaK	

Match informations

SpotID	ImageID

Information Console SpotList

Extension des interfaces

PARIS: Proteomic Analysis and Resources Indexation System [migadi connected on paris2_dev]

File Tools Preferences Configuration Help

Navigation

Zoom

no image.

Zoom Control

Data

Spot	Protein	Gel	Agmial
A	B		
Species	Staphylococcus aureus		
Strain	Mu50		
Genome_Accession	BA000017_GR		
ORF_tag	SAV1580		
Gene_Name	dnaK		
Position	complement(1689926..1691...		
Product	DnaK protein		
Protein_Accession			
Protein_Name	SaMu50#1#BA000017#4343		
Fonction			
Annotation_Status	Confirmed		

SaMu50#1#BA000017#4343 Agmial link

Gel

MOSAIC annotation

Data segmentation rearrangements

Staphylococcus aureus Mu50 (BA000017_GR)

Search

available releases: release5.1

available comparisons: 550

A	B
aligner	MGA2
nb_genomes	2
accessions	BA000017_GR, CP000703...

Mosaic link

gene name: dnaK

gene position: 1689926..1691758

downstream position: -1000

upstream position: +1000

view result

Result

segment_type	Backbone Segment (BS)
insertion	false
inversion	false
translocation	false

get alignment

get segmentation get rearrangement

Informations

Cons

Extension des interfaces

PARIS: Proteomic Analysis and Resources Indexation System [migadi connected on paris2_dev]

File Tools Preferences Configuration Help

Navigation
Zoom
no image.
Zoom Control

Data
Spot Protein Gel Agmial

A	B
Species	Staphylococcus aureus
Strain	Mu50
Genome_Accession	BA000017_GR
ORF_tag	SAV 1580
Gene_Name	dnaK
Position	complement(1689926..1691...
Product	DnaK protein
Protein_Accession	
Protein_Name	SaMu50#1#BA000017#4343
Fonction	
Annotation_Status	Confirmed

SaMu50#1#BA000017#4343 Agmial link

Gel

MOSAIC annotation

Data segmentation rearrangements
Staphylococcus aureus Mu50 (BA000017_GR)

Search

available releases: release5.1
available comparisons: 550

A	B
aligner	MGA2
nb_genomes	2
accessions	BA000017_GR, CP000703...

Mosaic link

gene name: dnaK
gene position: 1689926..1691758
downstream position: -1000
upstream position: +1000

view result

Result

segment_type	Backbone Segment (BS)
insertion	false
inversion	false
translocation	false

get alignment
get segmentation
get rearrangement

AGMIAL PARIS MOSAIC

Portail d'intégration

http://genome.jouy.inra.fr/migadi/demo

MIGADI AGMIAL MOSAIC PARIS

Web Services for Microbial Genome Annotation using Data Integration

This prototype relies on the integration of three existing applications via a web service architecture :

- [AGMIAL](#), an integrated system for bacterial genome annotation
- [MOSAIC](#), a database including systematic comparisons of closely related bacterial genomes
- [PARIS](#), a database including experimental data from 2D electrophoresis based proteomic analysis

Genome selection

select a reference strain :

By default, this study is on the genome length.
You can restrict your search by selecting a fragment (100..5000 for example).

..

Abims Station Biologique Roscoff
MIG MIA INRA

Terminé

MIGADI Result

Staphylococcus aureus Mu50

strain N315 vs strain Mu50 / ATCC 700699

gene name	dnaN
locus tag	SAV0002
gene position	2156..3289
gene product	DNA polymerase III beta chain
protein name	SaMu50#1#BA000017#2777
protein function	-
annotation status	Confirmed

MOSAIC comparison: MGA2 (MOSAIC)

segment type: Backbone Segment (BS)

rearrangement type: insertion = no, inversion = no, translocation = no

PARIS experiment: demo jobim

gels: SAureus_N315_PH4-7, SAureus_Mu50_PH4-7

AGMIAL MOSAIC PARIS

Applications links :

- [dnaN on AGMIAL](#) (login = agmialRead and password = agmialRead)
- [dnaN on MOSAIC alignment](#) (strain N315 vs strain Mu50 / ATCC 700699)
- [PARIS](#)

[Home Page](#)

Terminé

Conclusion

× Bilan

- Intégration de sources de données locales et distantes
- Web Services peu intrusifs et sans dépendance bloquante
- Enrichissement et évolution des applications

× Difficultés rencontrées

- Modélisation et mise au point d'ontologies nécessaires en amont (problèmes de sémantique posés par l'intégration)
- Gestion des données privées : homogénéisation des politiques de sécurité

× Perspectives

- Ajout de nouvelles sources de données : transcriptomique...
- Amélioration de la gestion de la sécurité dans un environnement distribué

Remerciements

× Financement

Programme des PlateFormes Technologiques du Vivant (2008-2010)

× Partenaires

-  , Station Biologique de Roscoff
Christophe CARON, Ludovic LEGRAND, Célia MICHOTEY
-  , INRA de Jouy-en-Josas
Hélène CHIAPELLO, Annie GENDRAULT, Jean-François GIBRAT,
Valentin LOUX, Véronique MARTIN
-  , INRA de Jouy-en-Josas
Sylvie HUET, Juhui WANG

× Collaborations

-  , INRA de Rennes
Stéphanie-Marie DEUTSCH, Hélène FALENTIN, Gwénaél JAN
-  , INRA de Jouy-en-Josas
•Véronique MONNET, Christophe GITTON

Web Services for Microbial Genome Annotation using Data Integration

Cécilia MICHOTÉY¹, Ludovic LEGRAND¹, Hélène CHIAPPELLO¹, Valérie LOUX¹, Anne GENDRAULT¹, Jean-François GIBRAT¹ and Christophe CARON²



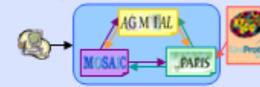
¹ Service Informatique et Génétique, FR2424 O IRIS-UPMC, Station Biologique de Roscoff, 29680 Roscoff, France
 Cécilia.michotey@roscoff.fr, Ludovic.legrand@roscoff.fr
² Unité Mathématique Informatique et Génome, UR 1077 MIRA, Domaine de l'Herminier, 78352 Jouy-en-Josas, France
 helene.chiapello@genotoul.fr, anne.gendault@jean-francois.gibrat@jouy.inra.fr



Le projet MIGADI (Microbial Genome Annotation using Data Integration) vise à développer une infrastructure bioinformatique innovante dédiée à l'annotation des génomes et des protéomes microbiens. La méthodologie proposée pour répondre à ce objectif est basée sur l'intégration virtuelle de données hétérogènes. Cette intégration repose essentiellement sur l'échange d'informations entre les environnements logiciels existants pour l'annotation des génomes microbiens, AGMIAL [1], et des outils plus ciblés sur deux domaines importants pour la microbiologie : la génomique comparée, via l'application MOSAIC [2] et la protéomique expérimentale, via l'application PARIS [3].

Architecture

L'interconnexion entre les applications est réalisée via une architecture de type Web Services, ce qui permet une implémentation à posteriori de composants hétérogènes.



Interfaces

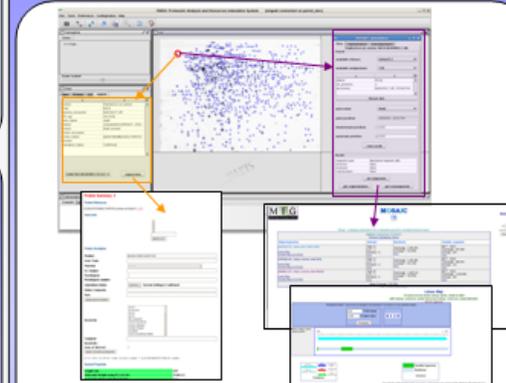


Figure 2 : hierarchie des applications AGMIAL, MOSAIC et PARIS autour celle dernière.

Méthodologie

Plusieurs scénarios impliquant les 3 composants ont été élaborés en collaboration avec des microbiologistes.

Question biologique (SCÉNARIO)	AGMIAL	MOSAIC	PARIS	Contraintes
Aide à l'annotation automatique d'un génome dans AGMIAL (séquences consommées/validées d'une souche dans AGMIAL)	C	F	-	- génome de référence existant dans l'espèce ? - gestion de la confidentialité
Aide à l'annotation de protéomes d'espèces dans AGMIAL (pour générer dans un job 20 de PARIS)	C	-	F	- gestion des génomes - gestion de la confidentialité
Aide à l'interprétation des protéomes expérimentaux dans PARIS (pour un 20 souche de la même espèce)	F	F	C	- gestion des génomes - gestion de la confidentialité

Tableau 1 : Exemples de scénarios utiles pour construire l'intégration des applications AGMIAL, MOSAIC et PARIS (C = Consommateur et F = Fournisseur).

Une API (Application Programming Interface) a été définie et implémentée entre les 3 composants logiciels du projet.



Figure 1 : Exemple de méthode implémentée. Utilisation de MOSAIC pour intégrer l'expression variable d'une protéine dans un gel 2D de PARIS.

Chaque application de web service utilise de données.

De nouvelles interfaces sont intégrées au sein de chaque application pour permettre l'ajout de données externalisées (cf. Figure 2). Un portail d'intégration a été créé (cf. Figure 3).

Les applications sont potentiellement liées.

L'intégration est testée avec des données de :
 - S. typhimurium (projet MISO et IN315 (données publiques provenant de GenBank et EpiSys))
 - Propionibacterium freudenreichii, souches CIRM1 et CIRM129 (données privées de STLO - INRA)



Figure 3 : Portail d'intégration (http://genome.jouy.inra.fr/migadi/demo).

Bilan

Les Web Services développés sont peu coûteux et facilement atomiques pour permettre des traitements complexes sans dépendance bloquante. Une réflexion complémentaire est nécessaire pour homogénéiser les processus d'authentification entre les applications et ainsi faciliter la gestion des données privées.

Conclusion et perspectives

L'intégration virtuelle réalisée autour des projets de recherche AGMIAL, PARIS et MOSAIC permet de valider et d'offrir de nouvelles perspectives à chaque application. L'intérêt de cette approche pour améliorer les processus d'annotation a été démontré par les premiers développeurs. De plus, le portail d'intégration permet l'accès à de nombreuses sources externes, répondant ainsi à un besoin croissant de croiser et lier des informations.

Un travail de modélisation et de mise au point d'ontologies est nécessaire et nous out pour évaluer les problèmes de sémantique posés par l'intégration. A terme, une extension du système est envisagée, notamment par l'intégration de nouvelles sources de données externes et de données de haut débit. L'amélioration de la gestion de la sécurité dans ce type d'environnement distribué sera également à prendre en compte.

Remerciements



References

- [1] K. Michotey, V. Loux, H. Chapelle, P. Nicolas, S. Chalhou, M. van der Guchte, S. Renault, S. Meguinard, J.-F. Gibrat. AGMIAL: implementing an annotation strategy for prokaryotic genomes as a distributed system. Nucleic Acids Research, 2006, 34, 2020-2045.
- [2] H. Chiapello, A. Gendault, C. Caron, J. Blain, M. H. Nicolas, M. G. H. Nicolas. MOSAIC: an online database dedicated to the comparative genomics of bacterial strains at the intra-species level. BMC Bioinformatics, 2008, 9:496.
- [3] Wang, C. Caron, M.-Y. Maucq, C. Gibot and A. Tribou. PARIS: a protein analysis and resource integration system. Bioinformatics, 2004, 20:159-166.

Poster n° 16

<http://genome.jouy.inra.fr/migadi/demo>

Portail d'intégration

<http://genome.jouy.inra.fr/migadi/demo>

The screenshot shows a Mozilla Firefox browser window displaying the MIGADI demo page. The browser's address bar shows the URL <http://genome.jouy.inra.fr/migadi/demo>. The page features a header with the MIGADI logo on the left and the logos for AGMIAL, MOSAIC, and PARIS on the right. The main heading is "Web Services for Microbial Genome Annotation using Data Integration". Below this, a paragraph explains that the prototype integrates three existing applications: AGMIAL, MOSAIC, and PARIS. A central section titled "Genome selection" contains a dropdown menu for "select a reference strain" with "Staphylococcus aureus Mu50" selected. Below the dropdown, text states "By default, this study is on the genome length. You can restrict your search by selecting a fragment (100..5000 for example)." and includes two input fields for "100" and "5000" separated by a double-dot range operator. An "Envoyer" button is positioned below the input fields. The footer of the page includes logos for Abims, Station Biologique Roscoff, MIG, MIA, and INRA, along with the text "Terminé".

MIGADI demo - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

<http://genome.jouy.inra.fr/migadi/demo> Google

Les plus visités Débuter avec Firefox À la une

MIGADI demo

MIGADI AGMIAL MOSAIC PARIS

Web Services for Microbial Genome Annotation using Data Integration

This prototype relies on the integration of three existing applications via a web service architecture :

- [AGMIAL](#), an integrated system for bacterial genome annotation
- [MOSAIC](#), a database including systematic comparisons of closely related bacterial genomes
- [PARIS](#), a database including experimental data from 2D electrophoresis based proteomic analysis

Genome selection

select a reference strain :

By default, this study is on the genome length.
You can restrict your search by selecting a fragment (100..5000 for example).

..

Abims Station Biologique Roscoff MIG MIA INRA

Terminé