



Étude Et Comparaison De Cartes Métaboliques De Centaines De Souches Bactériennes

Vers La Prédiction De Cocktails Produisant Des Composés Cibles ?

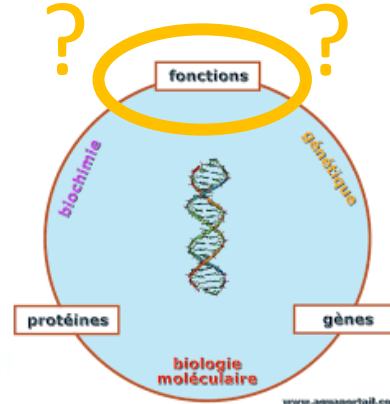
Anne SIEGEL – CNRS - Laboratoire IRISA - Equipe Dyliss
(et Hélène Falentin - INRAE – Labo STLO)



EXPLOITER DES DONNEES GENOMIQUES EN TERMES METABOLIQUES ?

Transformer l'information en fonctions : chercher une aiguille dans une botte de foin

- De dizaines à plusieurs milliers de génomes
- couvrant des **centaines d'ordres** taxonomiques
- codant pour un **centaine de millions** de séquences de **protéines**.

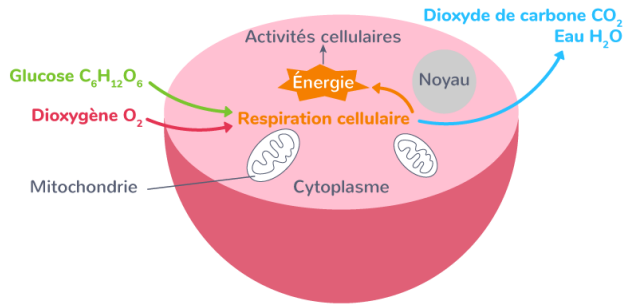


FOCUS SUR L'ECHELLE DU METABOLISME

Voies métaboliques: ensemble de **réactions chimiques** permettant de **convertir un composé chimique en un autre à travers des transformations successives**, catalysées par des enzymes.

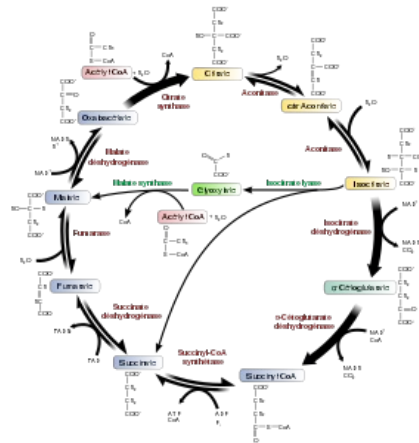
Fonction métabolique

La respiration cellulaire, une voie de production d'énergie



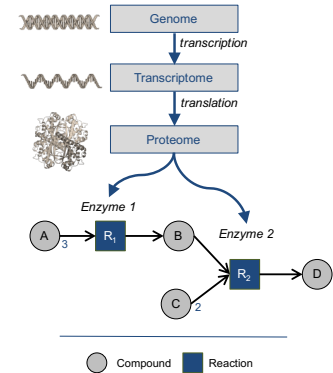
<https://www.kartable.fr>

Acteurs de cette fonction



https://fr.wikipedia.org/wiki/Cycle_du_glyoxylate

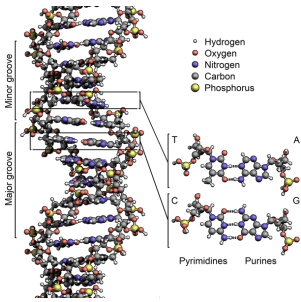
Du gène à la fonction



@C. Frioux

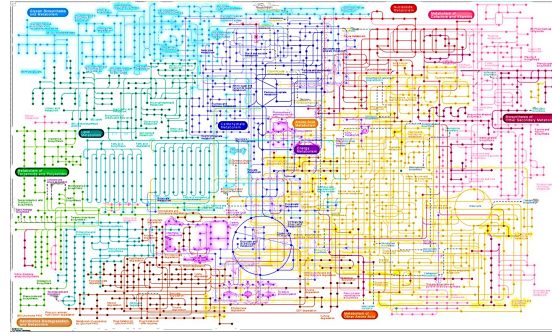
Les fonctions métaboliques sont contrôlées par des gènes

EXPLOITATION DU GENOME A L'ECHELLE DU METABOLISME



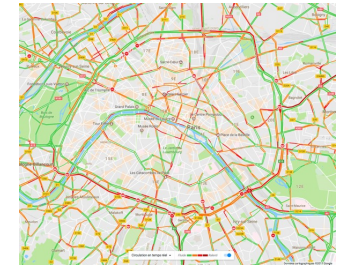
<https://fr.wikipedia.org>

Transformation du **génomé annoté** en carte métabolique



Source : KEGG

Analyse des flux: produire un maximum de composés sans provoquer d'embouteillages ?



insuline

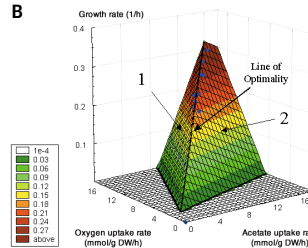
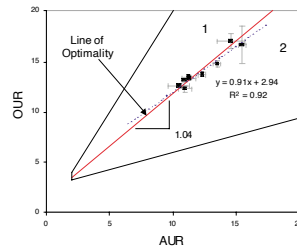
vitamines

cosmétiques

Acides aminés

ethanol

Génie génétique (CRISPR) A pour optimiser la production de molécules



Source : Edwards, Ibarra Palsson, nature biotech. 2001

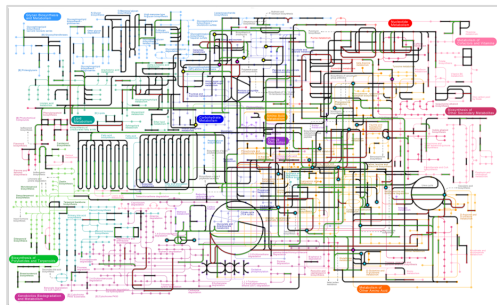
→ Systèmes dynamiques non identifiables
 → Principe thermodynamique & entropique du métabolisme
 → Recherche opérationnelle, raisonnement

Le schéma est valable pour les bactéries, levures et microalgues (unicellulaires) qui ont des cartes de bonne qualité

ESPECES « NON-MODELES » : EXEMPLE DES MACRO-ALGUES

Les cartes sont incomplètes

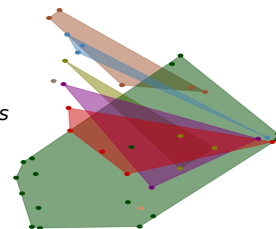
- on ne connaît que très partiellement les fonctions des protéines
- Il manque des voies de synthèse complètes dans les bases de connaissances



(Tapia et al, 2016)

Les cartes dépendent fortement des annotations et des technologies de séquençage

- Diatoms
- Green algae
- Haptophytes
- Nannochloropsis
- Brown algae
- Opisthokonts
- Red algae
- Glaucophytes
- Cryptophytes



Les réseaux métaboliques reconstruits pour une famille de 36 macro-algues n'ont aucune cohérence phylogénétique

(Belcour et al, Genome Research)



MGnify
@MGnifyDB

Using their fast ESMFold's language model this provides a structure for almost every cluster representative in the protein database, of which just under half of the proteins currently have an unknown functional annotation. #protein #darkmatter

La carte reconstruite à partir du génome d'E. siliculosus ne pouvait produire in silico aucun des composés mesurés expérimentalement

(PIA IDEALG, Prigent et al, 2015)

Chaque lacune dans une carte est un potentiel de découverte scientifique et de valorisation

COMPLÉTER LES CARTES... OU NON...

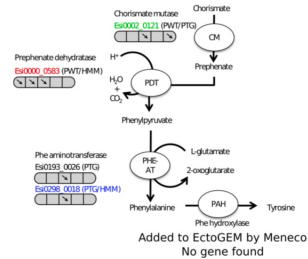
Méthode(s) in silico pour boucher les trous dans les cartes

(Logiciels: meneco, miscoto, AuReMe, AuCoMe...)

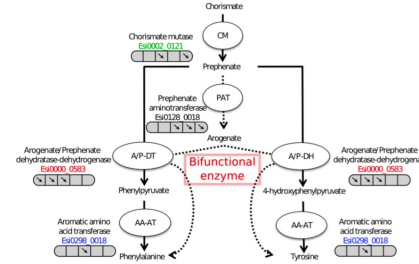
Synthèse de la tyrosine chez *E. siliculosus* : gène absent mais compréhension du rôle bifonctionnel d'une enzyme

(Prigent et al, 2014)

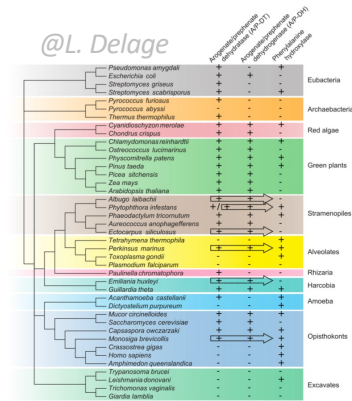
Proposed after automatic workflow



Proposed after manual curation



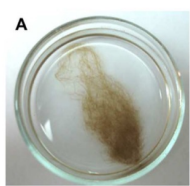
@L. Delage



Arrows: bifunctional enzymes

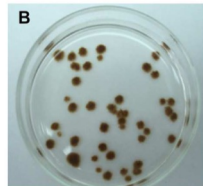
Voie de synthèse de la coenzyme A : nécessite de la beta-alanine et du pantoate

- L'algue sait produire le pantoate... mais n'a pas le gène prédit in silico pour produire la beta-alanine
- Ce gène est présent chez une bactérie symbiotique rescapée de la phase d'axénisation → Holobionte !



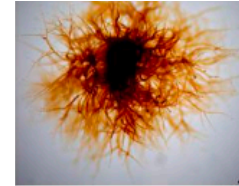
Sans microbiote
Condition axénique

(Tapia et al, 2016)



Ajout de 3 bactéries
prédites in silico

(Frioux et al, 2018)
(Burgunter-Delamare, 2020)



Identification de candidats (inexact) pour faire progresser la connaissance

SORTIR DU MAUVAIS SORT DE LA DIMENSIONALITE ?

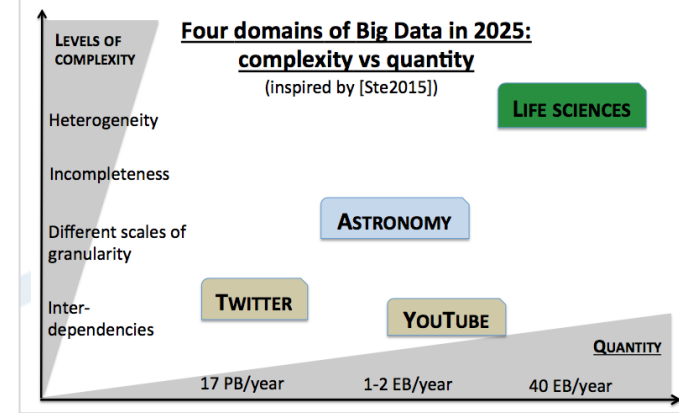
- Les données en sciences du vivants sont connues pour être les pires qui soient
- C'est la diversité des informations qui permet d'identifier des nouvelles voies de productions de composés

Principaux enseignements appris jusqu'à maintenant

- **Les petits détails font toute la différence du métabolisme**
- Ces petites différences sont impossibles à identifier automatiquement
- **L'informatique est nécessaire pour analyser les données, mais pas suffisante pour arriver à la découverte de connaissance.**

Une place de l'informatique/IA pour l'exploration de ces données.

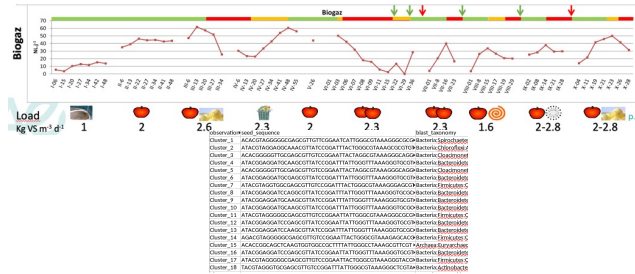
- **Qui** : who is there
- **Quoi** : who does what
- **Comment (bio)** : what is the role of each species
- **Comment (données)** : comment exporter de l'information intéressante des masses d'analyses



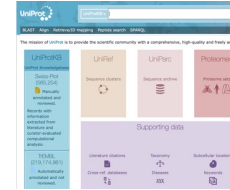
Data nightmare in life sciences

QUI ? WHO IS THERE

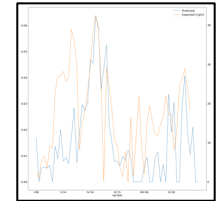
Des données de diversité (16S, taxons) aux fonctions



Metagenomic sequencing



Knowledge engineering
& sequence analysis
(Esmecata, Belcour et al, in progress)

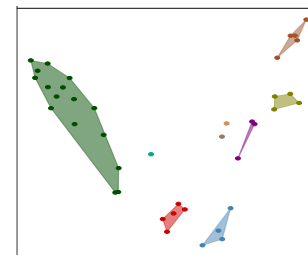
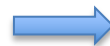
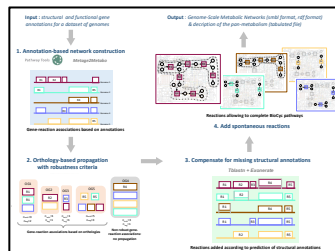
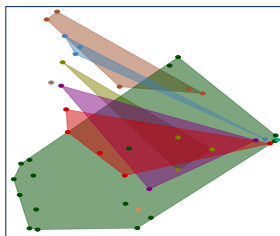


Functions

Extraire des connaissances sur les taxons/espèces et les croiser avec des alignements de séquences

QUOI : WHO DOES WHAT

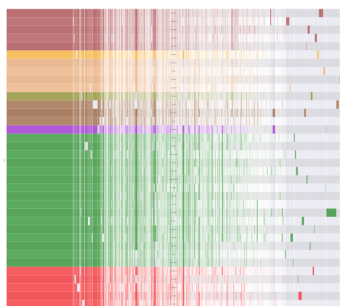
- Diatoms
- Green algae
- Haptophytes
- *Nannochloropsis*
- Brown algae
- Opisthochonts
- Red algae
- Glaucophytes
- Cryptophytes



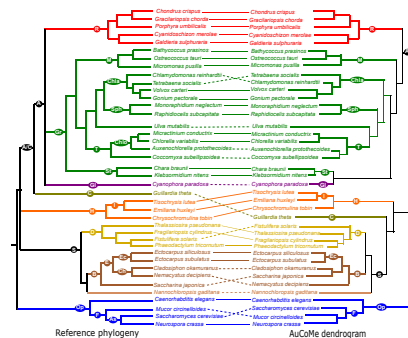
Distances entre les réseaux métaboliques associés à une famille d'algues

Gestion des métadonnées hétérogènes et comparaison/alignements de séquences
AuCoMe (*Belcour et al, Genome research*)

Réseaux comparables



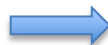
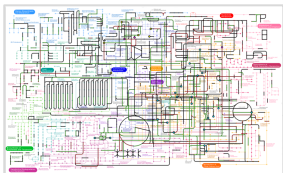
Identification de fonctions spécifiques



Des génomes hétérogènes aux cartes comparables
Besoin d'uniformisation des cartes pour pouvoir identifier des spécificités

COMMENT : WHAT IS THE ROLE OF EACH SPECIES

Compléter des cartes



Continuous framework (ODEs)

Dynamics $\frac{dM(t)}{dt} = Sv(M(t))$

Steady-state $\frac{dM(t)}{dt} = 0$

Side-constraints *Biomass optimisation
Flux upper/lower values*

Resolution framework *Linear (MILP) solving*

Discrete framework (Boolean)

$$\phi(\bar{m}) = \bigvee_{r \in R | m \in \text{product}(r)} \bigwedge_{s \in \text{substrate}(r)} \bar{s}$$

$$\phi(\bar{m}) = \bar{m} \quad \forall m \in M$$

$$\phi(\bar{m}) = 1 \quad \forall m \in \text{medium}$$

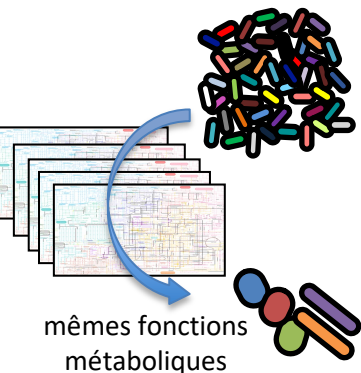
Minimal number of activated nodes

Logical programming with appropriate negation model (ASP)



Impact du microbiome intestinal sur la production de lipides (Belcour, Frioux et al, elife)

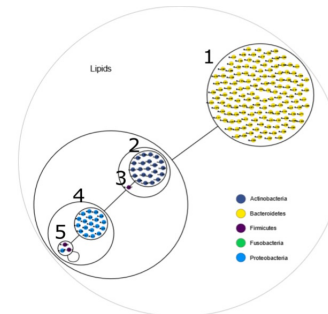
Réduire des consortia



mêmes fonctions métaboliques

Abstraction booléenne de la dynamique continue
Problèmes d'optimisation combinatoire complexes
(meneco, miscoto, metage2Metabo)

Associer des fonctions à des interactions d'intérêt (échanges, gap filling)



- Reduced communities :
- 1 Bacteroidetes from ①
 - 1 Actinobacteria from ②
 - Firmicute ③
 - 1 Proteobacteria from ④
 - All bacteria of ⑤

Visualization of metabolic complementarity from complete genomes.

STOCKER, PUIS EXPLORER LES DONNEES ET INFORMATIONS

Data, knowledge and workflow semantic engineering

L'informatique permet de répondre à des questions

Qui → Esmecata

Quoi → Mpwt, AuReMe, AuCoMe

Comment → Metage2Metabo

Workflow complexes qui combinent

- Des connaissances dans des bases de données
- Des génomes multiples
- Des alignments et comparaisons de séquences
- Des méthodes informatiques spécifiques

Gestion des données, formats, pipelines et métadonnées

→ Format « padmet » pour intégrer et exploiter les métadonnées biologiques et les métadonnées de pipeline

Vues multiples basées sur des wiki

The screenshot shows a multi-view wiki interface for pathway information. It includes several panels:

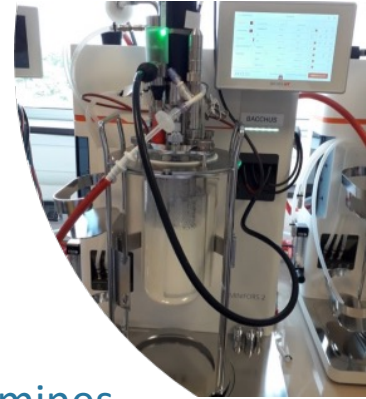
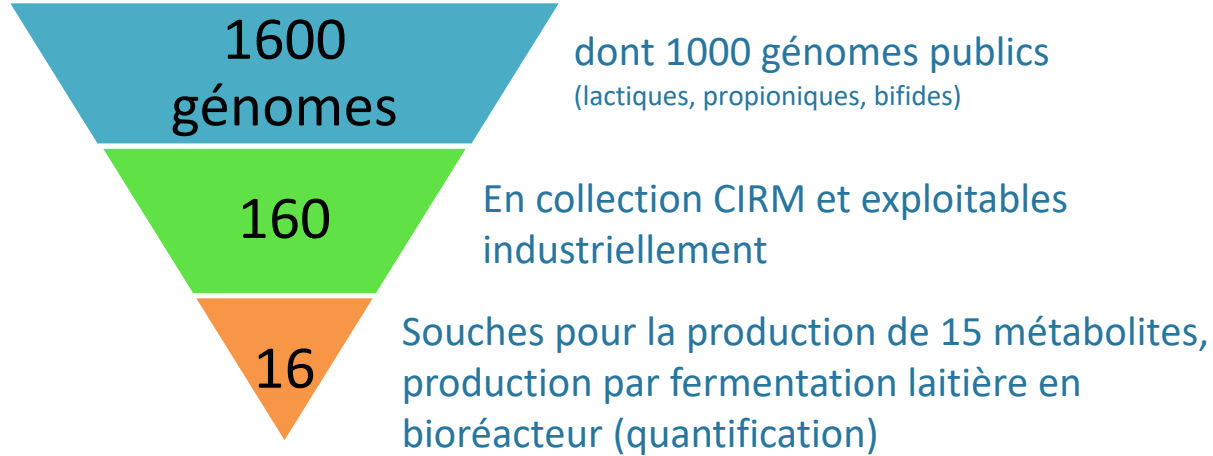
- Main page navigation panel:** Contains links for Semantic search, Categories (Reactions, Genes, Metabolites, Pathways), and Origins of reactions (EC number, Gene Ontology, etc.).
- Category table:** A table listing pathways with columns for Name, Number of reactions found, and Number of reactions NOT found.
- Pathway information:** Details for pathway PWY6614, including description, reactions found in EcoCyc, and external links.
- Reaction information:** Details for reaction DIHYDROFOLATESYNTH-RXN, including description, metabolites, and external links.
- Metabolite information:** Details for metabolite DIHYDROFOLATE, including description, reactions, and external links.
- External links:** A section for additional resources like METACYC.
- Main page search panel:** A search bar for finding specific pathways or metabolites.

Création d'entrepôts de données pour créer et exécuter des requêtes

The screenshot shows the Ask! query interface, which is used for creating and executing queries. It includes a search bar, a list of query results, and a table of query results. The table has columns for query name, query text, and query results.

The screenshot shows a network graph visualization, likely representing a metabolic pathway or a complex network of interactions. The graph consists of nodes (represented by colored circles) and edges (represented by lines connecting the nodes). The nodes are arranged in a hierarchical or branching structure, with some nodes highlighted in red. The interface includes a search bar and a table of query results at the bottom.

RESULTAT : PRODUCTION DE 15 METABOLITES EN SANTE DU LAIT



→ acides gras à courtes chaînes, neurotransmetteur, acides aminés, vitamines



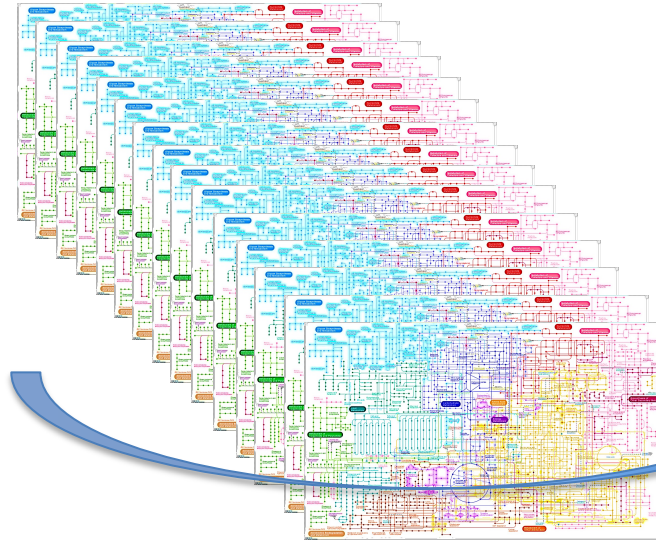
METHODE: CRIBLAGE IN SILICO POUR LES PRODUITS LAITIERS FERMENTES

Genre	Espèce	Nombre de souches par genre	Nombre de souches total par espèce
Bifidobacterium	breve	211	66
	longum		145
Lactobacillus	acidophilus	740	31
	alimentarius		2
	amylophilus		11
	brevis		46
	buchneri		16
	casei		9
	conviformis		6
	crispatus		70
	curvatus		20
	delbrueckii		68
	dioilovorans		1
	fermentum		25
	gallinarum		2
	gasseri		21
	helveticus		113
	hiigardii		9
	johnsonii		37
	kefiranoferiens		5
	kefir		4
	paracasei		35
	pentosus		39
plantarum	28		
reuteri	19		
rhamnosus	94		
sakei	25		
salivarius	4		
Lactococcus	lactis	104	104
	citreum		25
Leuconostoc	lactis	68	6
	mesenteroides		37
Pediococcus	acidilactici	77	35
	pentosaceus		39
Propionibacterium	freudenreichii	57	57
	thermophilus		113
Streptococcus	thermophilus	38	113
	ulyanov		10
Lactobacillae	total = 38		10
	total = 8		1180

1,600 genomes (8 genres, 38 espèces)

disponibles dans les bases de données publiques

ou dans les collections du CIRM-BIA



26 métabolites cibles

- Acide lactique
- GABA,
- Spermidine,
- Vitamines...

1,600 cartes métaboliques à construire et à analyser en fonction de leur capacité de production de métabolites cibles !

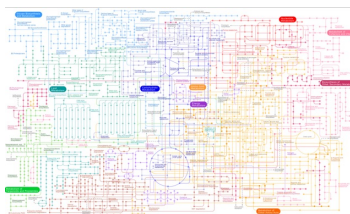
APPLICATION DES PRINCIPES DE DATA ENGINEERING



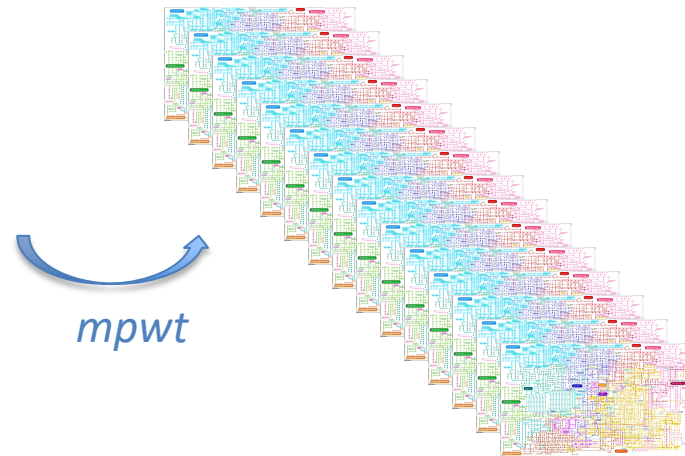
Genomes



Functional annotation
genomes



Large scale metabolic map



mpwt

Application to hundreds
of genomes

Prokka, eggno mapper

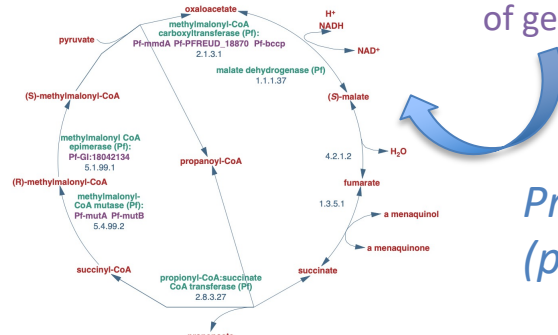
Pathwaytools



Askomics

Integration in a database

Resultat_B91_Pourcentage_de_complétion %	Souche_Label %	Resultat_B6_351_Pourcentage_de_complétion %	Resultat_B6
90	Lb_buchneri_CHR2253_P	85.89	non
90	SL3TH_CIRM_772_C	85.89	non
90	SL3TH_CIRM_961_C	85.89	non



Prolipipe (padmet)

Focus on specific pathways
Comparison on strain capabilities

INTERFACE NUMERIQUE POUR EXPLORER LA BASE DE DONNEES DE RESULTATS

Quelles sont les souches parmi les 1,600 qui ont le matériel génétique suffisant pour produire les vitamines B6 et B9 ? **REPONSE : 80 souches différentes**



ADCLY

Nombre de reaction corrigees

=

Nombre de reactions possedees

=

Nombre de reactions totales dans la voie

=

Pourcentage de completion

> 80

ADCLY-RXN-7

True

False

Run & preview

Resultat_B91_Pourcentage_de_completion ↑↓	Souche1_Label ↑↓	Resultat_B6_151_Pourcentage_de_completion ↑↓	Resultat_B6
90	Lb_buchneri_CNRZ2253_P	88.89	non
90	ST_STH_CIRM_772_C	88.89	non
90	ST_STH_CIRM_961_C	88.89	non

```
PREFIX : <http://askomics.org/data/>
PREFIX askomics: <http://askomics.org/internal/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX faldo: <http://biohackathon.org/resource/faldo/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```
SELECT DISTINCT ?Resultat_B91_Pourcentage_de_completion ?Souche1_Label ?Resultat_B6_151_Pourcentage_de_completion ?Resultat_B6_11_Souche_dependant
?Resultat_B91_Souche_dependant
WHERE {
  ?Resultat_B91_uri <http://askomics.org/data/PossedePar> ?Souche18_uri .
  ?Resultat_B6_151_uri <http://askomics.org/data/PossedePar> ?Souche18_uri .
  ?Resultat_B91_uri rdfs:type <http://askomics.org/data/Resultat20B9> .
  ?Resultat_B91_uri <http://askomics.org/data/Pourcentage%20de%20completion> ?Resultat_B91_Pourcentage_de_completion .
  ?Souche18_uri rdfs:type <http://askomics.org/data/Souche> .
  ?Souche18_uri rdfs:label ?Souche1_Label .
  ?Resultat_B6_151_uri rdfs:type <http://askomics.org/data/Resultat%20B6-1> .
  ?Resultat_B6_151_uri <http://askomics.org/data/Pourcentage%20de%20completion> ?Resultat_B6_151_Pourcentage_de_completion .
  ?Resultat_B6_151_uri <http://askomics.org/data/Souche%20dependant> ?Resultat_B6_151_Souche_dependantCategory .
  ?Resultat_B6_151_Souche_dependantCategory rdfs:label ?Resultat_B6_11_Souche_dependant .
  ?Resultat_B91_uri <http://askomics.org/data/Souche%20dependant> ?Resultat_B91_Souche_dependantCategory .
  ?Resultat_B91_Souche_dependantCategory rdfs:label ?Resultat_B91_Souche_dependant .

  FILTER ( ?Resultat_B91_Pourcentage_de_completion > 80 ) .
  FILTER ( ?Resultat_B6_151_Pourcentage_de_completion > 80 ) .
}
```

Les résultats des analyses sur les 1380 souches sont intégrés dans une base de données
Qui peut être explorée à l'aide de systèmes graphiques interactifs

EXPERIMENTATIONS

Résultat des analyses

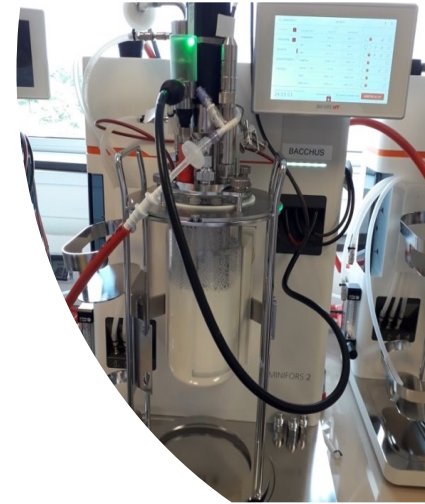
- La production de certains composés dépend des différentes souches
- Quelques cocktails de souches ont les capacités enzymatiques maximales
→ Candidats naturels pour des expérimentations !

WARNING : entre la théorie (numérique) et la pratique (biologie), il y a tout un univers !!!

- La bioinformatique n'est pas là pour remplacer la biologie mais pour guider l'expérimentation

STLO/INRAE (Hélène Falentin) → culture des cocktails de souches en bioréacteurs

- Difficulté à doser les composés pour valider la présence de composés
- Prestations de différents partenaires pour avoir toutes les informations nécessaires.
- Impact du milieu de culture et des modes de dosages.



15 voies de synthèses validées expérimentalement sur les souches sélectionnées.

DYNAMIQUE INTERDISCIPLINAIRE

Conception des consortia microbiens dirigés pour la production de matrices laitières fermentées ?

(1) Plateforme numérique pour le criblage in silico de plusieurs centaines de génomes de souches

- Intégration et comparaison de connaissances inexploitées sur la diversité des souches (génomés)
- Mise à disposition des résultats sous la forme d'une base de données partagée qui peut être explorée interactivement

(2) Dispositifs expérimentaux avancés pour tester les prédictions de cocktails de souches

- Cultures en bioréacteurs
- Mesures des composés cibles en collaboration avec différents partenaires

Déclinaisons possibles

- Analyse de la complémentarité métabolique des souches pour multi-fonctionnaliser
- Utilisation accrue de la biodiversité
- Choisir la matière première en fonction de la fonctionnalité visée

L'IA (science/gestion des données et connaissances) permet de
générer des candidats

Remerciements

Team Dyliss@IRISA

- C. Frioux (now @Inria Bordeaux)
- S. Prigent (now @Inra)
- A. Belcour
- M. Aite
- M. Chevallier
- J. Got
- C. Trottier

STLO/Inrae

- H. Falentin
- C. Raphalen
- Y. Tirlet
- Equipe MicroBio (Eric Guédon)
- CIRM-BIA (Florence Valence)

Station Biologique Roscoff: G. Markov, S. Dittami, C. Leblanc, P. Potin, B. Burgunger (...)



UMR

IRISA



STLO

Bactéries d'Intérêt Alimentaire