

CREATION D'UNE INTERFACE D'ANNOTATION METABOLIQUE COLLABORATIVE

Floréal Cabanettes

Master 2 Bioinformatique et Biologie des Systèmes

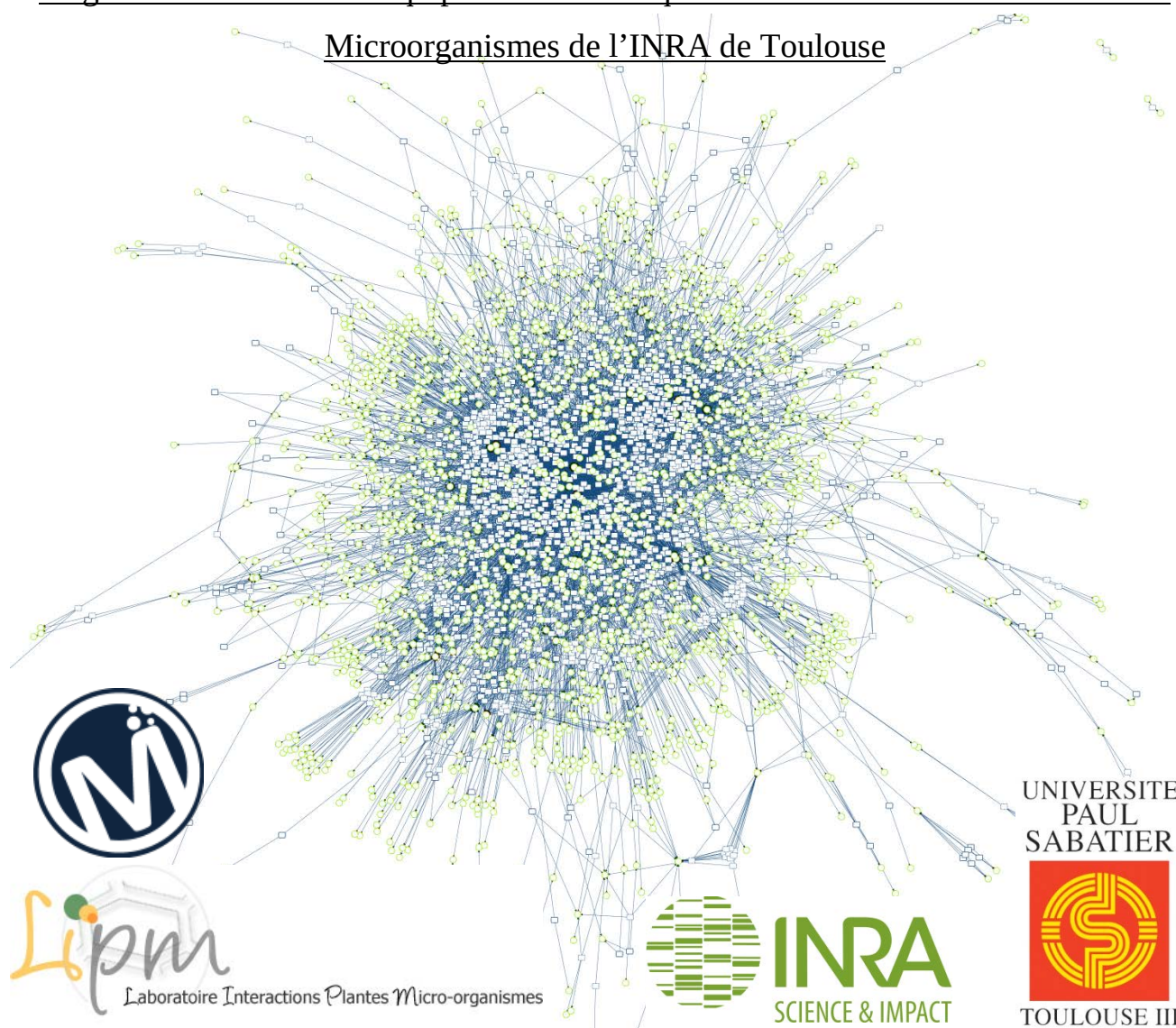
Encadrants :

01/06/2015 – Stage du 05/01 au 05/07/2015

- **Ludovic Cottret (LIPM, INRA Toulouse)**
- **Céline Lavire (Université, INRA Lyon)**

Stage réalisé au sein de l'équipe Bioinformatique du Laboratoire Interaction Plantes

Microorganismes de l'INRA de Toulouse



Laboratoire Interactions Plantes Micro-organismes



INRA
SCIENCE & IMPACT

UNIVERSITE
PAUL
SABATIER



TOULOUSE III

REMERCIEMENTS

Je remercie mes encadrants, *Ludovic Cottret* et *Céline Lavire*, pour m'avoir accepté pour ce stage dans lequel j'ai pris plaisir à participer et durant lequel j'ai beaucoup appris.

Je remercie tout particulièrement *Ludovic Cottret*, pour m'avoir encadré pendant ces 6 mois. Pour tous ses nombreux conseils, son aide très précieuse et sa patience.

Je voudrais également remercier *Marine Veyssière* pour sa compagnie dans le même bureau pendant les premiers mois du stage.

Bien sûr, un grand merci tous les membres de l'équipe *Bioinformatique* du LIPM, soit pour ceux non encore cités : *Sophie Siguenza*, *Érika Sallet*, *Jérôme Gouzy*, *Ludovic Legrand*, *Sébastien Carrere*, *Aziz Moussa*, *Soon Tantolin* et *Isabelle Gairin*, pour toutes les discussions que l'on a pu avoir et qui ont fait le « charme » et la bonne ambiance de ce stage.

Je tiens aussi à remercier l'équipe de *MetExplore* de Saint-Martin du Touch avec qui j'ai pu interagir, et m'enrichir : *Fabien Jourdan*, *Florence Vinson*, *Nathalie Poupin*, *Clément Frainay* ... Un remerciement spécial à *Benjamin Merlet* avec qui j'ai pu travailler pour certaines parties de mon stage, et un grand merci à *Maxime Chazalviel* pour son aide très précieuse lors de la semaine de code. Un merci à toute l'équipe de *MetExplore* présente lors de cette semaine de code.

Aussi, je remercie grandement mon rapporteur *Cédric Cabau* d'avoir accepté et pris le temps de lire ce présent rapport.

Je remercie enfin par avance tous les participants au *Jamboree*, pour tout ce qu'ils vont m'apporter lors de ces deux journées.

SOMMAIRE

Résumé.....	1
Abstract	1
Définitions.....	2
Introduction.....	3
Matériel & Méthodes	8
A. Présentation de l'Interface de <i>MetExplore</i>	8
B. Stratégie d'annotation du réseau métabolique	8
C. Architecture : outils informatiques utilisés	10
D. Stratégie de développement logiciel	11
E. Visualisation de l'association GPR	12
F. Gestion des droits dans <i>MetExplore</i>	13
Résultats	14
A. Panneau Utilisateur	14
1. La TODO List	15
2. L'historique.....	16
3. La liste des Projets de l'utilisateur	17
B. Panneau <i>Projet</i>	18
1. TODO List et Historique.....	19
2. Commentaires	19
C. Fiche Voie métabolique.....	21
D. Fiche Réaction.....	22
E. Panneau votes	23
F. Visualisation de l'association Gène-Protéine-Réaction (GPR) dans la fiche réaction.....	25
G. « <i>complétude</i> » des voies métaboliques	26
H. Export de tout le réseau au format Excel	28
Bilan et Perspectives	29
1. Favoriser l'annotation collaborative	30
2. Droits utilisateurs	30
3. Création d'outils d'évaluation du réseau métabolique.....	30
4. Favoriser le partage d'opinions.....	31
5. Renforcer l'annotation pour tous	31
Conclusion et perspectives.....	32
Bibliographie.....	33
Annexes.....	34
Figures supplémentaires.....	34
Description détaillé des principaux outils de reconstruction de la littérature	44

TABLE DES ILLUSTRATIONS

Figure 1. Définition d'un réseau métabolique.....	3
Figure 2. Organisation des données sous MetExplore	9
Figure 3. Schéma d'un MVC.....	10
Figure 4. Stratégie de développement d'une interface dans MetExplore.....	12
Figure 5. Chargement des données de l'association GPR.....	13
Figure 6. Panneau Utilisateur de MetExplore	14
Figure 7. Formulaire d'ajout d'une nouvelle action à la TODO List.....	15
Figure 8. Grille Historique de l'interface Utilisateur	16
Figure 9. Détails de la modification d'une réaction dans l'historique	17
Figure 10. Formulaire de création d'un nouveau Projet	18
Figure 11. Panneau Projet de MetExplore	19
Figure 12. Grille des commentaires de l'interface Projet	20
Figure 13. Formulaire d'ajout d'un nouveau commentaire	20
Figure 14. Formulaire d'ajout d'une pièce jointe à un commentaire	21
Figure 15. Fiche d'une Voie Métabolique	22
Figure 16. Fiche d'une Réaction	23
Figure 17. Panneau Votes d'une Réaction	24
Figure 18. Détails des votants d'une réaction	24
Figure 19. Ajout de la colonne résumant les votes des utilisateurs	25
Figure 20. Visualisation GPR réalisées à partir de 2 réactions	26
Figure 21. Affichage de la complétude des voies métaboliques dans la grille Pathways de l'onglet Network Data	27
Figure 22. Résumé de la complétude des voies métaboliques d'une BioSource	28
 Tableau 1. Comparaison de MetExplore + MetExplore Annotation aux principaux autres outils de reconstruction métabolique.....	 7
Tableau 2. Droits d'un utilisateur à une BioSource ou à un Projet.....	13

RESUME

Le processus de reconstruction métabolique consiste à définir la liste des réactions biochimiques intervenant dans le métabolisme d'un organisme. Afin de faciliter, d'accélérer et d'améliorer ce processus, nous avons complété le serveur Web *MetExplore* (Cottret *et al*, 2010) avec de nouvelles fonctionnalités de partage et d'amélioration de l'annotation des réseaux. Désormais, les utilisateurs enregistrés peuvent créer un projet et le partager avec d'autres utilisateurs afin que ceux-ci puissent également contribuer à l'annotation. A chaque projet peuvent être associées plusieurs reconstructions, ce qui facilite la propagation des annotations. Le propriétaire d'un projet définit des droits différents (lecture, écriture, ...) sur les réseaux du projet pour chaque participant. Nous avons également développé des outils d'évaluation du réseau métabolique, comme la visualisation des relations Gène-Protéine-Réaction ou des outils d'évaluation des voies métaboliques, dans le but de mettre en lumière des anomalies présentes dans ce réseau. Dans un processus d'annotation collaborative, nous avons mis en place également un système de vote pour définir la confiance que l'on a dans les entités du réseau. Par exemple, l'utilisateur peut voter pour la présence ou l'absence d'une réaction, ou, simplement signaler que certains de ses attributs sont erronés. Dans ce cas, il peut accompagner son vote d'un commentaire et d'un fichier attaché justifiant celui-ci. La décision finale revient alors au responsable du projet qui peut comparer les votes des différents participants du projet dans le but d'obtenir une meilleure reconstruction.

ABSTRACT

Metabolic network reconstruction consists in defining the list of the biochemical reactions involved in the metabolism of an organism. In order to facilitate, accelerate and improve this process, we have enhanced the *MetExplore* web server (Cottret *et al*, 2010) with new features of metabolic network sharing and curation. Registered users can now create a project and share it with other *MetExplore* users so that they can also contribute to the curation. Each project can contain several metabolic reconstructions, facilitating the propagation of the annotations. The owner of the project defines the rights of the users (read, write ...) on the metabolic reconstructions. We have also developed some evaluation tools of the metabolic network, like the Gene-Protein-Reaction association visualization or evaluation pathways tools, with the aim of highlighting some errors on the network. Lastly, to facilitate the collaborative annotation, we set a vote system. For instance, *MetExplore* users can vote for the presence or absence of a reaction, or can point out some erroneous attributes. They can also add comments, with an option to attach a file to support their inference. This enables the project owner to make a final decision based on the votes and comments in order to better the quality of reconstruction.

DEFINITIONS

Annotation : consiste à définir les composants d'une entité. Dans le cadre de l'annotation métabolique, c'est définir les éléments du réseau métabolique (réactions, métabolites, ...).

Association GPR : association gène-protéine-réaction : lien qui relie une réaction aux protéines qui la catalysent et aux gènes qui codent ces protéines.

BioSource : ensemble de tous les objets (Compartiments, Voies métaboliques, Réactions, Enzymes, Protéines et Gènes) qui composent une reconstruction métabolique.

Complétude d'une voie métabolique : pourcentage de réactions de la voie métabolique catalysées par une enzyme dans le réseau par rapport au nombre total de réactions qu'elle contient.

Objet / Entité : dans ce rapport, nous parlons d'objets ou d'entités pour désigner un composant d'une reconstruction métabolique. Ainsi, un objet peut désigner une voie métabolique, une réaction, un métabolite, une enzyme, une protéine ou un gène.

SBML : Systems Biology Markup Language, format d'échange standard d'annotation de réseau métabolique.

INTRODUCTION

L'information génétique, codée dans le génome, est transcrite en ARNs messagers, eux-mêmes traduits en protéines. Ces protéines peuvent être des enzymes catalysant une ou plusieurs réactions. Ces réactions transforment des métabolites, appelés substrats, en d'autres métabolites appelés produits (**Figure 1**). L'ensemble de ces réactions, reliées entre elles par les molécules qu'elles consomment et produisent, constituent le réseau métabolique de la cellule. Nous pouvons ajouter à ce réseau les relations Gène-Protéine-Réaction (GPR) qui permettent de retrouver pour chacune de ces entités (gène, protéine, réaction) les entités associées. Ce réseau, complexe, décrit le comportement physiologique de la cellule. Il varie en fonction du type cellulaire chez les organismes pluricellulaires, et de l'environnement de la cellule, notamment chez les bactéries.

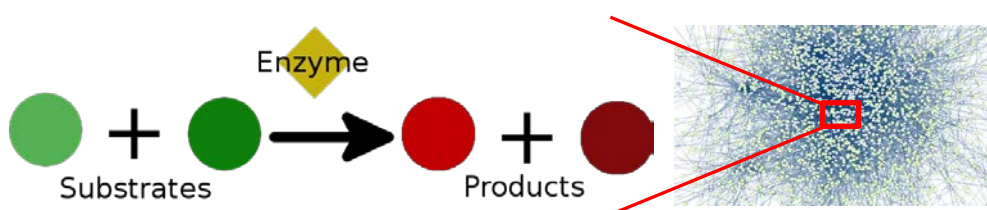


Figure 1. Définition d'un réseau métabolique. *Un réseau métabolique est constitué de réactions reliées entre elles par les métabolites (substrats et produits) qu'elles consomment ou qu'elles produisent.*

Ainsi, la modélisation de ce réseau permet d'étudier *in silico* la croissance cellulaire dans différentes conditions de milieu, afin par exemple de définir les conditions environnementales optimales à la croissance cellulaire, en comparant ce réseau métabolique dans différentes conditions environnementales. Il est également possible de réaliser des knock-out (KO) de gènes afin de déterminer *in silico* les gènes essentiels de la cellule. Enfin, grâce aux relations GPR du réseau métabolique, il est possible de visualiser les produits d'un ensemble de gènes sur le réseau. Ainsi, on peut visualiser le résultat d'une expérience de transcriptomique, ou d'une autre expérience -omique, sur le réseau. Aussi, grâce à la structure topologique du réseau, il est possible de trouver le plus court chemin qui relie deux métabolites ou deux réactions.

C'est dans cette optique que les chercheurs travaillent depuis plusieurs années sur la reconstruction de réseaux métaboliques. Elle consiste à combiner différentes sources de données – biochimiques, génétiques et génomiques (BiGG) notamment – pour reconstruire le réseau métabolique sous la forme finale d'un modèle mathématique, compréhensible pour l'ordinateur (Thiele & Palsson, 2010a). Cette reconstruction consiste à établir la liste des réactions, auxquelles sont associées ensuite les relations GPR.

Il est très difficile d'obtenir une reconstruction métabolique de qualité. La qualité d'une reconstruction dépend grandement de la quantité d'informations disponible pour l'organisme étudié, car c'est depuis ces informations que la reconstruction sera faite.

Malgré les avancées récentes, la reconstruction métabolique reste un travail long, de 6 mois pour des génomes bactériens bien connus et étudiés, de taille moyenne, à plus de 2 ans pour le génome humain par exemple, selon un protocole proposé récemment (Thiele & Palsson, 2010a). Ce protocole s'attelle à définir l'ensemble des étapes nécessaires, et de proposer pour chaque étape une manière de procéder. Il existe des reconstructions rapides, réalisées complètement automatiquement (dans les bases de données BioCyc et KEGG par exemple), mais la communauté scientifique s'accorde à dire que ces reconstructions sont trop imprécises car elles contiennent beaucoup d'erreurs (Swainston *et al*, 2011). Des caractéristiques spécifiques d'un organisme, tels que les substrats, l'utilisation de cofacteurs par des enzymes, ou encore le pH intracellulaire et le sens des réactions sont un véritable frein à l'automatisation du processus. La plupart des étapes doivent donc encore se faire manuellement, alors que seules certaines étapes peuvent être automatisées. Un challenge de la bioinformatique est de proposer des outils facilitant la reconstruction ou la correction manuelle des réseaux métaboliques.

D'après Thiele *et al.* (2010), la reconstruction métabolique est décomposée en 4 grandes étapes. La première est la reconstruction d'un *brouillon*. C'est pour ainsi dire la seule étape que l'on peut faire de manière automatique. Ce brouillon est réalisé à partir de l'annotation du génome. Depuis ces annotations (*E.C. numbers*, nom des enzymes, *GO [Gene Ontology] terms*), les fonctions métaboliques de la cellule sont extraites. Grâce à elles, la liste des réactions candidates est extraite depuis des bases de données métaboliques comme KEGG ou BioCyc. Ces bases de données peuvent être utilisées pour récupérer pour chaque gène les protéines et réactions associées, c'est-à-dire les relations GPR. La liste des réactions candidates obtenue ne sera pas nécessairement correcte et complète, car elle contiendra nombre de faux positifs qu'il faudra ensuite éliminer, ou omettra certaines réactions qu'il faudra ajouter. D'autres réactions seront erronées et il faudra les corriger.

Cependant, la reconstruction automatique, en plus des erreurs qu'elle contient, ne contient pas d'informations sur la réversibilité des réactions et leur localisation dans les compartiments cellulaires. Tout ceci va être ajouté lors de la seconde étape, correspondant à l'amélioration manuelle de la reconstruction métabolique. C'est une étape importante mais très longue durant laquelle chaque réaction de la reconstruction métabolique va être réévaluée manuellement et corrigée si nécessaire. Ainsi, pour chacune d'entre elles, il faut se poser deux questions :

- (1) Cette réaction existe-t-elle réellement et est-elle correcte ?
- (2) Y a-t-il une ou plusieurs réaction(s) manquante(s) pour connecter cette réaction au reste du réseau ?

Ainsi, les différentes réactions sont vérifiées individuellement en utilisant la littérature, ou les bases de données, spécifiques de l'organisme étudié. C'est d'autant plus important que les sources de don-

nées utilisées pour la reconstruction automatique ne sont pas toutes spécifiques de l'organisme étudié. Ensuite, l'appartenance de chaque réaction à une ou plusieurs voie(s) métabolique(s) est vérifiée pour faciliter l'interprétation du réseau, et les relations GPR sont également vérifiées.

La troisième étape correspond à la traduction de la reconstruction métabolique en modèle mathématique, compréhensible par l'ordinateur. Elle peut être automatisée. Des outils existent pour cela, comme la COBRA Toolbox (Becker *et al*, 2007). Ceci permet de pouvoir ensuite tester le réseau informatiquement, lors de la dernière étape, celle de l'évaluation du réseau.

Lors de la dernière étape d'évaluation du réseau, ce dernier sera testé par des outils permettant d'identifier des réactions manquantes qui déconnectent une partie de ce réseau du reste. Il faudra alors s'aider de bases de données de référence pour trouver un moyen de les trouver et de compléter le réseau. Enfin, il faudra tester si le système est capable de croître : le modèle est-il capable de produire chaque composé de la biomasse dans des conditions de milieu standard ? Certaines réactions sont-elles bloquées (ne peuvent conduire aucun flux dans aucune condition simulée) ? Il faudra également tester les (in)capacités connues de l'organisme dans le modèle. Enfin, il faudra évaluer quantitativement le taux de croissance : est-il trop lent ? Trop rapide ? Il faut corriger le réseau le cas échéant. Tout ceci se fait par la modélisation des flux qui permet de visualiser la répartition des flux dans le réseau. À l'inverse, la modélisation des graphes permet, elle, d'étudier la topologie du réseau, comme les chemins qui relient deux entités de ce réseau.

Ce stage s'inscrit dans le projet *Agromics*, qui a pour vocation d'étudier plusieurs espèces du genre *Agrobacterium* localisées dans des environnements très différents, et de visualiser l'impact de ces environnements sur le réseau métabolique. Lors de ce stage, nous nous intéresserons à la reconstruction métabolique d'*Agrobacterium tumefaciens*, une bactérie gram- retrouvée dans les sols et responsable de la galle du collet, une maladie qui infecte les végétaux, et principalement les dicotylédones dont font partie nombre de plantes d'intérêt agronomique. L'étude de cette bactérie représente donc un intérêt économique. Cette espèce constitue une espèce référence du genre. On pourra donc par la suite partir de cette reconstruction pour étudier celle d'espèces moins connues du genre.

Plusieurs outils sont disponibles actuellement pour réaliser une reconstruction automatique d'un brouillon à partir de l'annotation du génome. Nous pouvons citer KAAS (Moriya *et al*, 2007), utilisé pour créer les réseaux KEGG, *Pathway tools* (Karp *et al*, 2002) associé à la base de données BioCyc, ou encore la COBRA Toolbox (Becker *et al*, 2007) et la Raven Toolbox (Agren *et al*, 2013) tournant sous Matlab, ainsi que GEMSiRV (Liao *et al*, 2012) et Model Seed (Henry *et al*, 2010).

Dans le cadre du projet d'annotation du réseau d'*Agrobacterium tumefaciens*, nous ne souhaitons pas repartir de la séquence génomique pour reconstruire le réseau. À l'inverse, nous souhaitons par-

tir de réseaux existants, reconstruits automatiquement (KEGG, BioCyc et un réseau reconstruit via la *SuBliMinaL Toolbox* (Swainston *et al*, 2011)). Ces réseaux comportant encore des erreurs, il est nécessaire de les corriger manuellement en accord avec les données de la littérature.

Nous avons donc besoin d'un outil permettant l'édition d'un réseau déjà reconstruit. Le nombre d'outils permettant cela est plus restreint. *Pathway tools* (Karp *et al*, 2002) permet l'édition du réseau, mais ne propose pas d'importer ou d'exporter le réseau en SBML, un format d'échange standard de réseaux métaboliques entre plusieurs applications, ou au format Excel. *RbioNet* (Thorleifsson & Thiele, 2011) permet également l'édition de réseau métabolique, tout en proposant l'import ou l'export de réseau au format SBML et Excel. *GEMSiRV* (Liao *et al*, 2012), un autre outil permettant l'édition d'un réseau préexistant, grâce à l'import depuis le format SBML ou Excel, ne propose pas d'export. Il propose cependant de réaliser des simulations sur le réseau comme des Analyses de flux. La *Raven Toolbox* (Agren *et al*, 2013) propose les mêmes chose, et également l'export en SBML ou Excel du réseau, ainsi que l'export sous *Cytoscape*. Enfin, *SuBliMinaL Toolbox* (Swainston *et al*, 2011) permet de fusionner des reconstructions KEGG et BioCyc, et/ou des reconstructions depuis un fichier SBML, et de générer un modèle réunissant l'information de plusieurs sources de données.

Aucun des outils cités ci-dessus ne propose une interface d'annotation collaborative, à l'exception de *Pathway tools*, mais elle reste cependant très limitée. Le projet *Agromics* impliquant de nombreux collaborateurs, nous avons besoin d'un outil collaboratif pour effectuer cette tâche. Par ailleurs, les collaborateurs du projet viennent de différents horizons (chimistes, biologistes, bioinformaticiens, ...). Tous n'ont donc pas la même maîtrise des outils informatiques. C'est pourquoi nous avons besoin d'une interface qui s'adapte à tous. Tout ceci ne se retrouve pas dans un même outil pré-existant.

C'est pourquoi nous avons développé notre propre interface d'annotation, *MetExplore Annotation*. Cette interface permet de modifier, corriger les réactions directement dans la base de données. Mais elle permet maintenant également, grâce au travail que j'ai effectué durant de stage, d'indiquer son opinion pour chaque composant du réseau. Ceci se fait grâce à un système de vote que j'ai développé durant mon stage, associé à un système de commentaires permettant de commenter chaque entité du réseau. J'ai également mis en place une interface utilisateur qui permet d'avoir une vue d'ensemble du travail effectué et à faire, grâce à une gestion de l'historique et un système de *TODO list*. Et j'ai également créé un système de *Projets* dans lequel plusieurs personnes participent, comme le *Projet Agrobacterium tumefaciens*. À l'intérieur d'un même *Projet*, il y a donc plusieurs collaborateurs qui travaillent à corriger un ensemble de réseaux communs à tous, où chacun a des droits bien définis selon son rôle dans le *Projet*. Enfin, j'ai permis l'export du réseau au format Excel qui autorise ceux qui sont plus à l'aise avec ce format d'apporter leurs modifications.

	MetExplore + M. Annotation	KAAS	Pathway tools	rBioNet	GEMSiRV	Pathway Booster	Raven Toolbox	Model SEED	SuBliMinaL Toolbox
Reconstruction automatique	Non	Oui	Oui	Non	Oui	Non	Oui	Oui	Non
Corrections manuelles	Oui	Non	Oui	Oui	Oui	Non	Oui	Non	Oui
Import/Export SBML	Oui	Non	Non	Oui	Import : oui Export : non	Non	Oui	Non	Oui
Import/Export Excel	Import : non* Export : réseau complet	Non	Non	Oui	Import : oui Export : non	Non	Oui	Non	Non
Import direct depuis KEGG	Oui	Non	Non	Non	Non	Non	Non	Non	Oui
Import direct depuis BioCyc	Non	Non	Non	Non	Non	Non	Non	Non	Oui
Annotation collaborative**	Oui	Non	Oui	Non	Non	Non	Non	Non	Non
Gestion de projets	Oui	Non	Non	Non	Non	Non	Non	Non	Non
Simulations sur le réseau	Oui	Non	Non	Non	Oui	Non	Oui	Oui	Non
Outils d'évaluation***	Oui	Non	Non	Non	Non	Non	Non	Non	Non
Visualisation du réseau	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non	Non
Mapping sur le réseau	Oui	Non	Oui	Non	Non	Non	Non	Non	Non
Comparaison de réseaux	Non	Non	Non	Non	Non	Oui	Non	Non	Non
Export dans Cytoscape	Oui	Non	Non	Non	Non	Non	Oui	Non	Non

* Bientôt disponible, ** Une annotation est collaborative si plusieurs personnes peuvent modifier le réseau depuis différents ordinateurs, *** Outils permettant d'évaluer les composants du réseau, comme un système de votes ou de commentaires.

Tableau 1. Comparaison de *MetExplore* + *MetExplore Annotation* aux principaux autres outils de reconstruction métabolique.

Nous avons choisi d'intégrer cette interface au sein du serveur *MetExplore* (Cottret *et al*, 2010). *MetExplore* possède une interface de visualisation avancée du réseau métabolique qui complète et aide l'annotation. Également, différentes analyses peuvent être réalisées, comme des analyses de flux, du mapping et leur visualisation sur le réseau. Ces analyses peuvent aider à la correction de l'annotation. C'est pourquoi nous avons intégré notre interface d'annotation, *MetExplore Annotation*, à *MetExplore*. Le **Tableau 1** compare les fonctionnalités de *MetExplore Annotation* aux principaux autres outils de reconstruction métaboliques existants.

Dans le cadre de ce stage, j'ai été accueilli par l'équipe Bioinformatique du LIPM (Laboratoire Interaction Plantes Micro-organismes) de l'INRA (Institut National de la Recherche Agronomique) d'Auzeville (Toulouse). Cette équipe, dirigée par Jérôme Gouzy, est composée de 3 ingénieurs de recherche et 2 ingénieurs d'études en bioinformatique, 2 assistants ingénieurs en Informatique et plusieurs CDD en bioinformatique. Mon travail m'a également amené à travailler avec l'équipe de *MetExplore*, composée de 11 personnes répartis sur les sites d'Auzeville et de Saint-Martin du Touch (Toxalim), ainsi qu'une personne de l'Université de Glasgow (Écosse).

MATERIEL & METHODES

A. PRESENTATION DE L'INTERFACE DE *METEXPLORE*

MetExplore est un serveur web permettant notamment de visualiser des réseaux métaboliques. Ainsi, l'interface est composée trois onglets : *Network Data*, contenant des grilles listant les différents composants du réseau : les *BioSources* – contenant chacune un réseau métabolique, les Compartiments cellulaires du réseau, les Voies métaboliques (*Pathways*), les Réactions, les Complexes enzymatiques, les Protéines et les Gènes (**Figure S10** en annexe) ; *Network Viz*, permettant de visualiser le réseau sous forme d'un graphe ; et *Network Curation* permettant de modifier des composants du réseau. À gauche de ces onglets, j'ai ajouté deux onglets permettant d'ouvrir les panneaux *Utilisateur* et *Projet*, décrits dans la suite du rapport.

B. STRATEGIE D'ANNOTATION DU RESEAU METABOLIQUE

Dans le cadre du projet *Agromics*, la correction de l'annotation métabolique de différentes espèces d'*Agrobacterium* impliquera plusieurs collaborateurs. Chacun aura un rôle différent dans ce travail. Certains s'attacheront par exemple davantage à la vérification des informations tandis que d'autres auront pour tâche la correction des données dans la base de données via *MetExplore*. Ainsi, les premiers auront uniquement besoin d'indiquer leurs opinions, via des votes et/ou des commentaires. Les seconds auront besoin d'avoir un droit en lecture/écriture afin de pouvoir appliquer les modifications suggérées par les premiers.

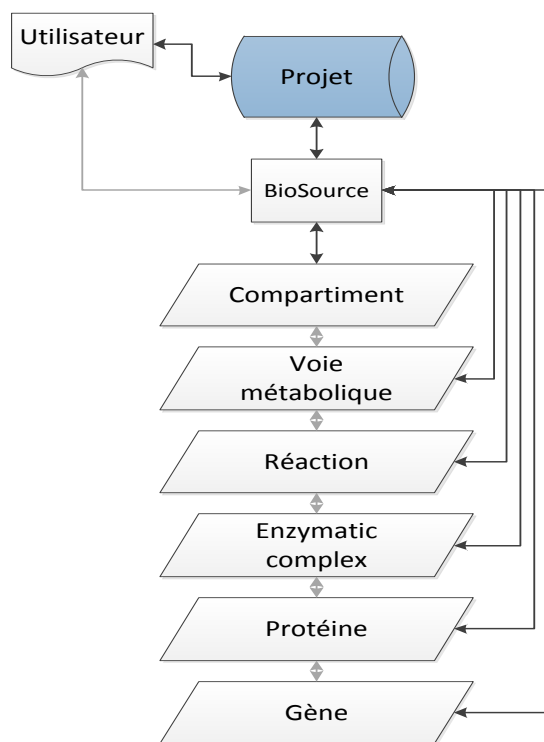


Figure 2. Organisation des données sous MetExplore. En gris, les relations secondaires.

désormais en moins de trois clics.

Ensuite, il manquait à *MetExplore* un moyen d'avoir une trace de toutes les modifications effectuées sur une ou plusieurs *BioSource(s)*. Pour pallier à ceci, j'ai mis en place un historique, qui enregistre dans la base de données l'ensemble des modifications qui sont apportées à une *BioSource* ou à l'un de ses composants.

Dans l'optique d'un travail collaboratif, il manquait également la possibilité pour les utilisateurs de planifier les tâches à effectuer directement dans l'interface de *MetExplore*. C'est pour répondre à ceci que j'ai créé une *TODO List*, permettant à chaque utilisateur d'ajouter une tâche pour lui ou un autre utilisateur du même *Projet*.

Afin de faciliter l'annotation, il était enfin nécessaire d'ajouter des outils de visualisation de l'annotation. C'est ce qui a été fait en créant les différentes fiches pour chaque composant du réseau métabolique. Également la visualisation des relations GPR a été mise en place dans la fiche Réaction dans cette optique, de même que les graphiques permettant de visualiser la complétude des voies métaboliques, c'est-à-dire le nombre de réactions qui possèdent au moins une enzyme sur le nombre total d'enzymes, pour chaque voie métabolique.

Tous ces outils permettront d'organiser les 16 et 17 juin un *Jamboree* (Thiele & Palsson, 2010b). Un *jamboree* consiste à réunir des spécialistes de différentes spécialités pour réaliser une tâche commune : la correction manuelle du réseau métabolique reconstruit automatiquement

Aussi, ce projet nécessitera d'utiliser plusieurs *BioSources*, contenant chacune un ensemble de compartiments cellulaires, de voies métaboliques, de réactions, de métabolites, d'enzymes, de protéines et de gènes. Chaque utilisateur devra avoir les mêmes droits sur toutes ces *BioSources*. Or jusqu'ici, dans la structure des données de *MetExplore*, chaque *BioSource* était directement reliée à l'utilisateur. Il a donc été décidé d'ajouter une nouvelle entité, le *Projet*, qui peut contenir plusieurs *BioSources*. Ainsi, les droits d'un utilisateur à un *Projet* s'appliquent à l'ensemble des *BioSources* du *Projet* (**Figure 2**). Ce *Projet* facilite également le partage d'un ensemble de *BioSource* à d'autres utilisateurs, puisque ceci peut se faire

d'*Agrobacterium tumefaciens* pour ce premier *Jamboree*. *MetExplore Annotation* a donc été conçu dans le but de faciliter l'organisation de ces deux jours, qui sera organisée par mon encadrant et moi-même.

C. ARCHITECTURE : OUTILS INFORMATIQUES UTILISES

Côté client, l'interface est intégralement codée en JavaScript, au travers de l'environnement ExtJS 4. ExtJS permet d'organiser le code sous forme de classes, et donc de faire de la programmation orientée objet en JavaScript. ExtJS est une structure permettant de réaliser des applications riches pour le Web, qui propose différentes interfaces intuitives qui maximisent l'interface homme-machine, telles que des grilles pour les données – permettant une représentation simple et claire des données, en permettant d'afficher ou non certaines colonnes, de regrouper des lignes en groupes ou encore de colorer certaines lignes – ou des graphiques. C'est pour cela qu'il a été utilisé pour coder *MetExplore*, afin de faciliter son développement. Ainsi, comme pour *MetExplore*, *MetExplore Annotation* utilise les objets déjà présents dans ExtJS pour ses interfaces.

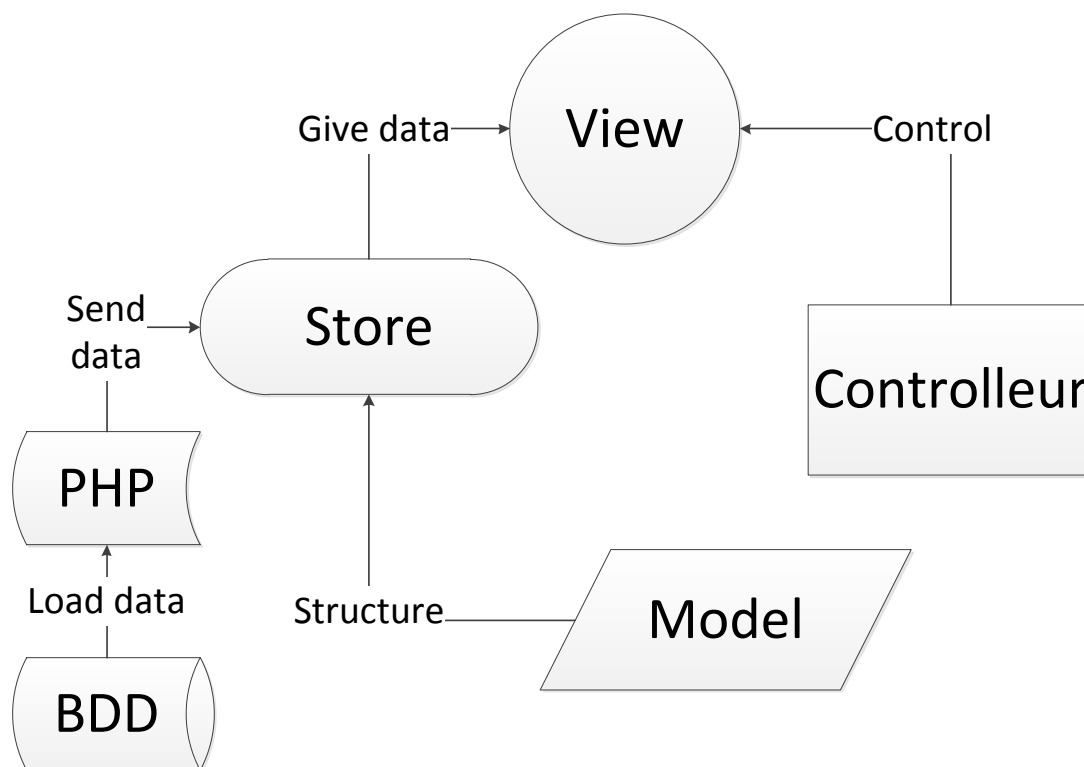


Figure 3. Schéma d'un MVC. Le Store contient les données de la Vue, chargées depuis la base de données sur le serveur grâce à un script PHP. La structure de ces données est définie dans le Modèle. Les actions de la Vue sont définies dans le contrôleur, qui la contrôle.

Dans le cadre du projet *MetExplore*, nous utilisons l'architecture Modèle-Vue-Contrôleur (MVC) pour créer nos classes ExtJS. Les données sont chargées depuis le serveur et stockées dans des classes appelées *Store*. Chaque *Store* contient une collection de *Modèles*, qui définissent un certain nombre d'attributs associés à un type de données. Ces attributs correspondent aux colonnes dans

une grille. Ainsi, un même *Modèle* peut être utilisé par plusieurs *Stores* sans redondance de code. Ces données sont affichées à l'utilisateur dans une interface utilisateur qui est définie dans une classe appelée *Vue*. Dans cette classe sont définis tous les composants affichés à l'écran. À certains de ces composants, comme des boutons, peuvent être associées des actions, qui sont définies dans une autre classe appelée *Contrôleur*. Le modèle MVC est résumé **Figure 3**.

Côté serveur, les données sont stockées dans une base de données MySQL. Ces données sont récupérées via des scripts PHP. Généralement, les classes ExtJS, notamment les *Stores*, font appel à ces PHP via des requêtes Ajax, et le script PHP renvoie à la classe ExtJS les données au format JSON.

Le code est stocké avec un système de versions sur un serveur SVN hébergé par *Sourceforge.net*. Ceci permet le développement logiciel collaboratif entre les différents développeurs de *MetExplore*. En effet, le serveur SVN permet de regrouper l'ensemble des modifications apportées par tous tout en gérant les conflits éventuels. Il contient également les différentes versions du code, ce qui permet de garder une trace de ce qui a été fait et de revenir si nécessaire à une version antérieure d'un ou plusieurs fichier(s).

D. STRATEGIE DE DEVELOPPEMENT LOGICIEL

Le développement de *MetExplore Annotation* s'inscrit dans le développement de *MetExplore*. Il a donc fallu respecter les standards établis. Dans un premier temps, le temps de tout bien maîtriser, mon travail s'est fait dans une branche à part du SVN, avant que je fusionne cette branche avec la branche principale après validation de mon travail par l'équipe de *MetExplore*.

Le développement de chaque interface du projet a suivi un chemin bien précis (**Figure 4**). Tout d'abord, il a fallu que je définisse en accord avec mon encadrant les objectifs de l'interface. Ensuite, j'ai créé un prototype non fonctionnel de l'interface, en proposant un agencement de ses différents composants, et en définissant le schéma des classes nécessaires et des tables MySQL nécessaires. Les **Figures S1 à S9** en annexe présentent les prototypes des panneaux *Utilisateur* et *Projet* tels qu'ils ont été conceptualisés et validés par l'équipe de *MetExplore* avant leur implémentation, ainsi que les schémas des classes nécessaires à leur implémentation. Ensuite, tout ceci est validé ou non par l'équipe de *MetExplore*. Si des modifications sont à faire, elles sont ajoutées au prototype de l'interface. Quand l'interface est validée par l'équipe, je peux alors l'implémenter dans *MetExplore*. L'interface est ensuite testée pour identifier les bugs et d'éventuels manques à cette interface, et une documentation est rédigée de même qu'un protocole de test permettant de s'assurer facilement et par l'ensemble de l'équipe de *MetExplore*, sans besoin de connaissances en programmation, que l'interface reste fonctionnelle malgré les modifications qui peuvent être apportées à *MetExplore* par la suite.

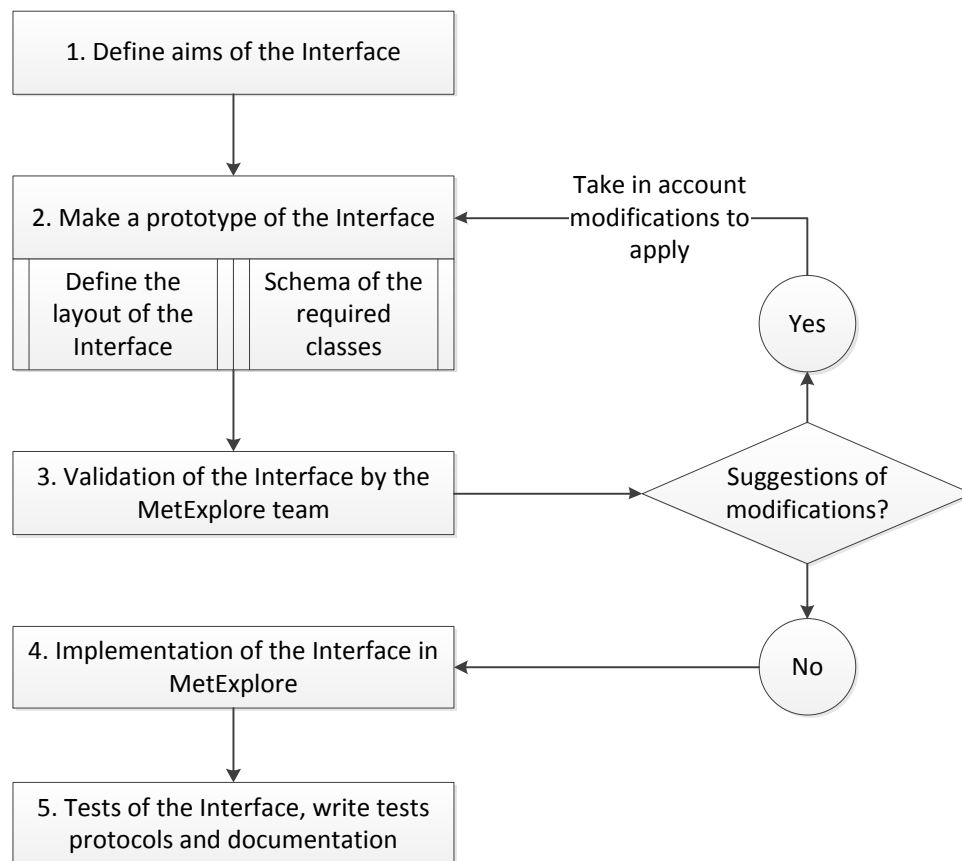


Figure 4. Stratégie de développement d'une interface dans *MetExplore*.

Le développement s'est donc fait en interaction étroite avec l'équipe de *MetExplore*. C'était d'autant plus le cas lors d'une *semaine de code* qui a été organisée, et qui a réuni toute l'équipe de *MetExplore* en un même lieu. La présence d'utilisateurs réels de *MetExplore* a permis d'obtenir des suggestions pour améliorer l'expérience utilisateur. Mais c'était également l'occasion de favoriser l'interaction entre tous les développeurs du projet. Ainsi, j'ai pu travailler en collaboration avec *Maxime Chazalviel*, un autre stagiaire qui travaille sur la visualisation du réseau dans *MetExplore*, afin de développer la visualisation GPR.

Aussi, afin de ne pas dupliquer le code, j'ai repris quand c'était possible le code déjà présent dans *MetExplore*, pour certaines fonctions, mais aussi pour les vues dont certaines sont *héritées* de vues déjà existantes dans *MetExplore*.

E. VISUALISATION DE L'ASSOCIATION GPR

La visualisation de l'association Gène-Protéine-Réaction (GPR) est réalisée en *d3.js* (Bostock *et al*, 2011), déjà utilisé pour la visualisation du réseau sous *MetExplore viz*, qui permet de créer une visualisation guidée par les données. Ainsi les données sont contenues dans une structure JSON. Nous voulons créer un graphe reliant les gènes, les protéines et les réactions associées. Pour se faire, les données doivent contenir un ensemble de nœuds correspondant à ces objets, et un ensemble de liens

indiquant comment les relier. Tout ceci est généré comme indiqué **Figure 5**. Ensuite, *d3.js* s'occupe de dessiner le graphe avec un algorithme de force qui permet de placer les nœuds automatiquement en regroupant les éléments très connectés entre eux, et en les éloignant des éléments moins connectés à eux. Ceci permet de former des groupes d'éléments connectés et de tracer automatiquement un affichage aussi clair que possible.

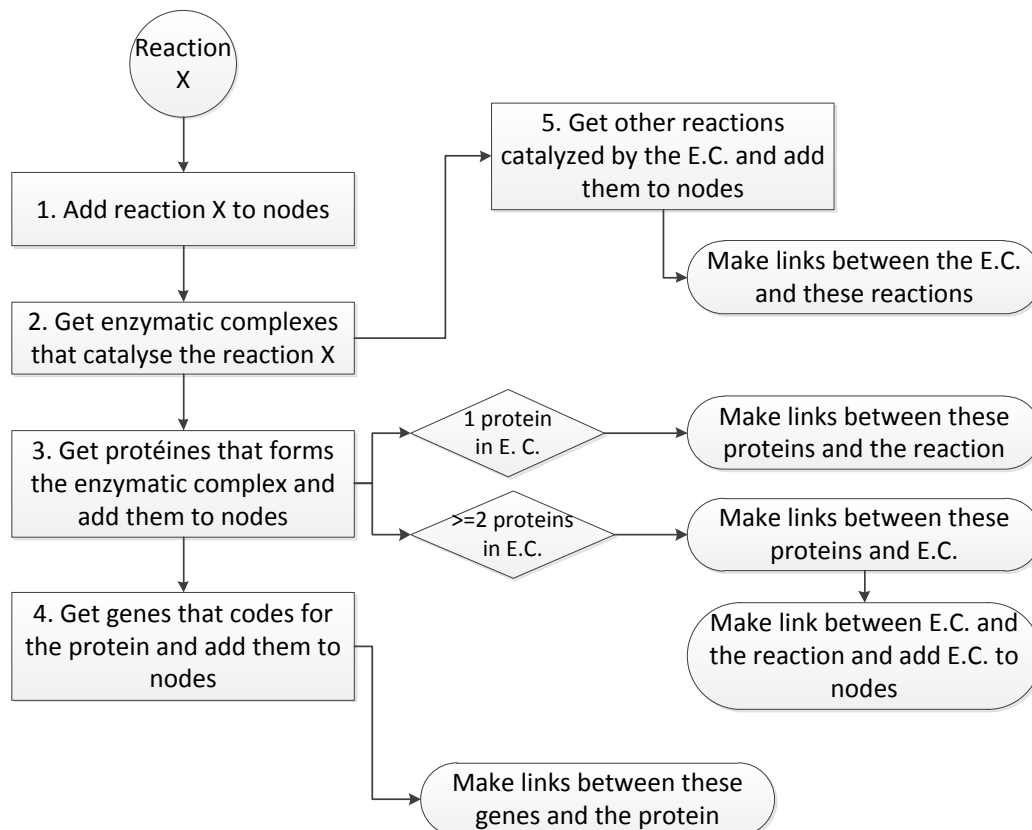


Figure 5. Chargement des données de l'association GPR. E.C. : enzymatic complex.

F. GESTION DES DROITS DANS *METEXPLORE*

Jusqu'ici, un utilisateur avait accès ou non à une *BioSource*. C'était binaire. Pour le partage de *BioSource* au travers ou non d'un *Projet*, il est nécessaire de définir différents droits aux utilisateurs y ayant accès. Ces droits sont résumés dans le **Tableau 2**. Ils permettent de donner des droits différents à des contributeurs d'un *Projet* ou d'une *BioSource* en fonction du rôle qu'ils y ont.

Nom	Autorisations
<u>Propriétaire</u>	L'utilisateur peut effectuer toutes les modifications.
<u>Lecture/Écriture</u>	L'utilisateur peut modifier les données mais ne peut pas modifier les utilisateurs du <i>Projet</i> ou de la <i>BioSource</i> , ni supprimer le <i>Projet</i> ou la <i>BioSource</i> .
<u>Annotateur</u>	L'utilisateur ne peut qu'ajouter des commentaires ou voter pour les objets du <i>BioSource</i> ou pour les objets des <i>BioSources</i> du <i>Projet</i> .
<u>Lecture seule</u>	L'utilisateur peut tout voir mais ne peut pas faire de modification.
<u>Accès refusé</u>	L'utilisateur n'a pas accès à la <i>BioSource</i> ou au <i>Projet</i> .

Tableau 2. Droits d'un utilisateur à une *BioSource* ou à un *Projet*. Classés du haut vers le bas dans le sens décroissant des privilèges utilisateur.

RESULTATS

Ma participation au développement de *MetExplore Annotation* a nécessité la création de nombreuses interfaces. Dans un effort de synthèse, nous vous décrivons ci-dessous les principales.

A. PANNEAU UTILISATEUR

Le panneau *Utilisateur* est pensé comme un panneau d'entrée dans *MetExplore*. Ainsi, depuis ce panneau, l'utilisateur a accès à un résumé du travail fait ou à faire dans les *Projets* auxquels il a accès. Ainsi, en haut du panneau apparaît la *TODO List* des actions faites ou à faire (**Figure 6**). En bas, dans différents onglets sont présents :

- La liste des *Projets* auxquels l'utilisateur a accès.
- La liste des *BioSources* privées auxquelles l'utilisateur a accès.
- L'historique de toutes les modifications apportées aux *BioSources* auxquelles l'utilisateur a accès.

En haut du panneau, est affiché le nom de l'utilisateur connecté, à côté un bouton pour se déconnecter et en bas un bouton pour modifier ses informations personnelles (**Figure 6**).

The screenshot displays the 'User profile' tab of the MetExplore interface. At the top right, the user 'agromics' is logged in, with an 'Edit profile' button. The 'TODO list' section contains a table of tasks:

Description	Project	User	Limit date	Status
Cure reactions of pathways with completeness less than 100%	Project Agrobacterium	agromics	2015-06-03	Not started
Remove pathways with no reactions in KEGG BioSource	Project Agrobacterium	Floréal Cabanettes	2015-06-01	Not started
Cure reactions of pathways with completeness less than 60%	Project Agrobacterium	agromics	2015-06-01	Not started
Cure reactions of pathways with completeness less than 30%	Project Agrobacterium	Cottret Ludovic	2015-05-29	In progress
Check proteins of BioCyc BioSource	Project Agrobacterium	agromics	2015-05-25	Done

Below the TODO list is a 'My projects' section with a table:

Name	Access	Last modification
Agrobacterium radiobacter	owner	2015-05-13
Agrobacterium rhizogenes		
Project Agrobacterium	owner	2015-05-27

At the bottom, there are navigation buttons: 'Add', 'Open', 'Unsubscribe', and 'Delete'.

Figure 6. Panneau *Utilisateur* de *MetExplore*. Les boutons de la barre d'outils de droite de la *TODO List* sont comme suit. Pour le groupe du haut, ils permettent de changer le statut de la *TODO List*. Le statut devient, du haut vers le bas : En cours, Non commencé, Terminé et Abandonné. Le groupe du bas permet, de haut en bas : d'ajouter une nouvelle action et de supprimer une ou plusieurs action(s).

1. LA TODO LIST

Objectifs

La *TODO List* regroupe l'ensemble des actions faites ou à faire de l'ensemble des utilisateurs pour tous les *Projets* auxquels l'utilisateur connecté participe.

Fonctionnement

La *TODO List* affiche par défaut uniquement les actions associées à l'utilisateur connecté. Mais il est possible de visualiser l'ensemble des actions de tous les utilisateurs en cliquant sur le bouton « *All* » en bas de la grille. On obtient ainsi une liste comme celle montrée **Figure 6**. Pour voir de nouveau que ses propres actions, il faut cliquer sur « *Personal* ».

Chaque action a une description, un *Projet* associé, un utilisateur associé, une date limite (échéance) et un statut. Le statut peut être : *pas commencé* (« Not started »), *en cours* (« In progress »), *terminé* (« Done ») ou *abandonné* (« Canceled »). Si l'échéance d'une action est dans moins de cinq jours (ou passée), et que le statut n'est pas *terminé* ou *abandonné*, l'action apparaît sur fond rouge pour indiquer qu'elle doit être terminée bientôt, comme la quatrième action en partant du haut sur la **Figure 6**.

Pour modifier le statut d'une ou plusieurs action(s), il suffit de le(s) sélectionner et de cliquer sur le bouton statut voulu, en haut dans la barre d'outils à droite de la grille (voir description de la **Figure 6**). Pour ajouter une nouvelle action, un bouton est présent dans la barre d'outils de droite. En cliquant dessus, un formulaire s'affiche alors permettant d'indiquer les informations de l'action à ajouter (**Figure 7**). L'utilisateur connecté est placé par défaut à l'ouverture du formulaire, et la liste des utilisateurs se met à jour en fonction du *Projet* sélectionné. Pour modifier une action, l'utilisateur doit double-cliquer dessus. Le formulaire montré **Figure 7** s'affiche alors, pré-rempli avec les valeurs d'origine de l'action. Pour supprimer une ou plusieurs action(s), l'utilisateur les sélectionne et clique sur le bouton correspondant dans la barre d'outils de droite. La modification ou la suppression d'une action n'est possible que si c'est sa propre action ou que l'on est un propriétaire du *Projet* associé.

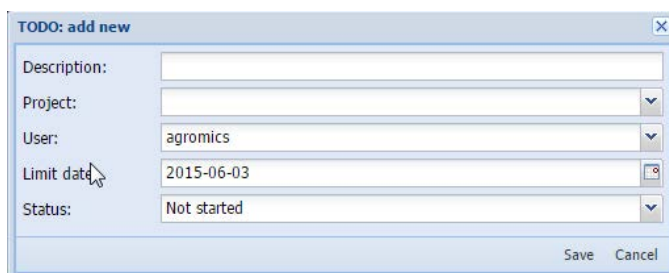


Figure 7. Formulaire d'ajout d'une nouvelle action à la *TODO List*.

2. L'HISTORIQUE

Objectifs

Il est important lorsque l'on travaille sur la reconstruction métabolique, comme pour beaucoup de travaux scientifiques, de garder une trace de ce que l'on fait. L'ajout d'un historique permet ici de visualiser facilement l'ensemble des modifications qui sont apportées à une *BioSource* ou à l'un de ses composants. Ceci peut également faciliter le travail lors de la publication de la reconstruction métabolique.

Fonctionnement

My projects		My BioSources		History	
Date	User	Project	BioSource	Action	
2015-05-27	agromics	Agrobacterium tumefaciens	Agrobacterium fabrum KEGG ...	Update pathway "atu02010"	...
2015-05-27	agromics	Agrobacterium tumefaciens	Agrobacterium fabrum KEGG ...	Add reaction R00136	...
2015-05-27	agromics	Agrobacterium tumefaciens	Agrobacterium fabrum KEGG ...	Update reaction R00135	...
2015-05-27	Floréal Cabanettes	Agrobacterium tumefaciens	Whole Genome Metabolism - ...	Update reaction R_bigq_quin_kt_out	...

Figure 8. Grille Historique de l'interface Utilisateur.

L'historique du panneau *Utilisateur* affiche, dans l'un des trois onglets du bas (**Figure 6**), l'historique associé à toutes les *BioSources* auxquelles l'utilisateur a accès, pour tous les utilisateurs. Comme pour la *TODO List*, au chargement de l'interface, l'historique n'affiche que les entrées associées à l'utilisateur connecté grâce à un filtre. Il suffit là encore de cliquer sur « *All* » dans la barre d'outils du bas pour afficher celles de tous les utilisateurs (**Figure 8**). Le bouton « *Perso-
nal* » permet de revenir à l'état initial.

Pour chaque entrée, l'historique affiche sa date, l'utilisateur ayant réalisé l'action, le *Projet* associé s'il y en a un, la *BioSource* associée et une description de l'action réalisée. Pour avoir plus d'informations quant à cette action, il suffit de cliquer sur l'icône  de la ligne concernée, dans la dernière colonne. Les détails s'affichent dans une nouvelle fenêtre. Dans le cas d'un ajout d'un objet, on peut voir les détails du nouvel objet. Dans le cas de la suppression d'un objet, on peut voir les détails de l'objet qui a été supprimé. Enfin, dans le cas de la mise à jour d'un objet, la fenêtre affiche l'ancienne version en haut, sur fond rouge, et la nouvelle version en bas, sur fond vert. Les champs modifiés apparaissent sur fond jaune (**Figure 9**). La génération se veut pour le moment générique au niveau du code, et n'est pas toujours très claire du point de vue utilisateur. Ces détails sont actuellement essentiellement faits pour aider les organisateurs du *Projet*, mais devront par la suite être améliorés pour être compréhensibles par tous.

History details					
Update reaction:					
	name	Biblio	EC	Product	Substrat
old	rn:R00135		3.4.11.5	[{"idMetabolite"...]	[{"idMetabolite"...]
new	proline iminope...		3.4.11.5	[{"idMetabolite"...]	[{"idMetabolite"...]

Close

Figure 9. Détails de la modification d'une réaction dans l'historique.

Au chargement de l'interface, l'historique affiche les entrées depuis la date de l'entrée la plus récente trouvée dans la base de données pour l'utilisateur connecté, jusqu'à 15 jours avant cette date. Il est possible de modifier les dates de début et de fin directement dans les champs de date présents dans la barre d'outils du bas de la grille. Dès la modification de ces dates, l'historique est rechargé. Il est également possible de naviguer dans l'historique en gardant l'intervalle de dates indiqué dans les champs. Ceci se fait grâce aux flèches de la barre d'outils du bas, de part et d'autre des champs de date. Celle de gauche permet de visualiser les entrées plus anciennes, et celle de droite les entrées plus récentes.

3. LA LISTE DES PROJETS DE L'UTILISATEUR

La liste des *Projets* auxquels a accès l'utilisateur apparaît dans l'un des trois onglets du bas du panneau *Utilisateur* (**Figure 6**). La liste est structurée en une grille à trois colonnes : le nom de chaque *Projet*, l'accès que l'utilisateur a à ce *Projet*, et sa date de dernière modification. Dans cette liste, les *Projets* dont l'utilisateur n'a pas encore accepté l'invitation apparaissent sur fond gris et seul le nom est visible. Ceux qui ont été modifiés par d'autres utilisateurs depuis la dernière ouverture du *Projet* par l'utilisateur courant apparaissent sur fond bleu.

Après avoir sélectionné un *Projet*, on peut l'ouvrir, via un menu contextuel, ou via le bouton de la barre d'outils du bas, ou en double-cliquant dessus. L'ouverture consiste à la création et l'affichage du panneau *Projet* (voir paragraphe B ci-dessous). L'utilisateur peut également se désinscrire du *Projet*. Enfin, l'utilisateur peut supprimer le *Projet* s'il en est un propriétaire. Toutes ces actions sont possibles via un menu contextuel, ou depuis les boutons de la barre d'outils du bas, et ne sont possibles que si l'utilisateur est un utilisateur actif du *Projet*. Dans le cas contraire, il doit d'abord accepter l'invitation au *Projet*, via un menu contextuel ou en double-cliquant dessus. Il peut aussi refuser l'invitation via ce menu contextuel.

L'utilisateur peut également créer un nouveau *Projet*, via le bouton « Add » de la barre d'outils du bas. Ceci provoque l'ouverture d'un formulaire dans une nouvelle fenêtre permettant d'entrer les informations du nouveau projet (**Figure 10**) : le nom, la description et les utilisateurs du *Projet*. Les

utilisateurs ajoutés, outre l'utilisateur courant, ajouté automatiquement, devront accepter l'invitation avant de faire réellement partie du *Projet*. En attendant leur acceptation, leur nom apparaîtra en italique dans la grille des *Utilisateurs*.

Name	Access
agromics	owner

Add new user

Username or e-mail:

Figure 10. Formulaire de création d'un nouveau *Projet*.

B. PANNEAU *PROJET*

Le panneau *Projet* a pour objectif de concentrer au sein d'une même interface l'ensemble des informations du *Projet* et du travail fait ou à faire pour ce *Projet*. Pour une meilleure expérience utilisateur, la structure de ce panneau est construite de la même manière que celle du panneau *Utilisateur*. Ainsi, on retrouve en haut la *TODO List* sur toute la largeur (**Figure 11**). En bas, dans les différents onglets sont présents :

- La liste des *BioSources* associées au *Projet*
- La liste des Commentaires associés au *Projet*
- L'historique de toutes les modifications apportées aux *BioSources* du *Projet*
- La description du *Projet*
- Les utilisateurs du *Projet*.

Dans l'en-tête du haut est affiché le titre du projet, à côté un bouton pour le fermer et en-dessous la date de création du *Projet* et un bouton pour modifier les détails du *Projet*, si l'utilisateur en est un propriétaire. Ce dernier ouvre le formulaire montré **Figure 10**, pré-rempli avec les données d'origines du *Projet*.

User profileProject DetailsNetwork DataNetwork VizNetwork Curation

Project Agrobacterium

Created 2015-05-12

Edit project

TODO list

Description	User	Limit date	Status
Cure reactions of pathways with completeness less than 100%	agromics	2015-06-03	Not started
Remove pathways with no reactions in KEGG BioSource	Floréal Cabanettes	2015-06-01	Not started
Cure reactions of pathways with completeness less than 60%	agromics	2015-06-01	Not started
Cure reactions of pathways with completeness less than 30%	Cottret Ludovic	2015-05-29	In progress
Check proteins of BioCyc BioSource	agromics	2015-05-25	Done

Personal

All

BioSources

Comments

History

Description

Users

ID	Name	Organism	Strain	Source Database	Database Type	Publication
2000	AGT5A	Agrobacterium tumefaciens	5A	Agromics	biocyc	
2001	AGRT5A	Agrobacterium tumefaciens	5A v2	Agromics	biocyc	
2595	Agrobacterium fabrum KEGG Genes Data...	Agrobacterium tumefaciens	C58	Kegg	Kegg	
3008	Whole Genome Metabolism - Agrobacteri...	Agrobacterium tumefaciens	C58	Subliminal...	SBML	Swainston...

Add BioSource to the project

Figure 11. Panneau *Projet* de *MetExplore*. Voir **Figure 6** pour la description des boutons de la *TODO List*.

1. TODO LIST ET HISTORIQUE

La *TODO List* et l'historique du *Projet* sont exactement les mêmes que pour le panneau *Utilisateur*, mais sont filtrés pour n'afficher que les entrées correspondant au *Projet* ouvert.

2. COMMENTAIRES

Objectifs

L'objectif est d'offrir une interface permettant d'ajouter des commentaires généraux à tout le *Projet*, en y associant si nécessaire des pièces jointes apportant un plus à l'information du commentaire, comme une illustration ou un texte rédigé à part.

Fonctionnement

Les commentaires font partie des onglets du bas de l'interface *Projet* (**Figure 11**). Les commentaires s'affichent dans une grille avec leur auteur, leur titre et leur nombre de pièces attachées (**Figure 12**). Une fois un commentaire sélectionné, on peut l'ouvrir, ou le supprimer, si l'on est l'auteur du commentaire ou qu'on est un propriétaire du *Projet* associé. Tout ceci se fait via un menu contextuel ou via les boutons de la barre d'outils du bas. L'ouverture affiche un formulaire dans une nouvelle fenêtre (**Figure 13**), pré-rempli avec les données du commentaire ouvert. On peut également créer un nouveau commentaire si on est au minimum un annotateur du *Projet*, en cliquant

sur le bouton correspondant dans la barre d'outils du bas. Ceci ouvre le même formulaire, mais vide comme sur la **Figure 13**.

BioSources Comments History Description Users			
	User	Title	Attachments
1	Floréal Cabanettes	New reconstruction available for <i>A. tumefaciens</i>	None
2	Floréal Cabanettes	Preparation for the Jamboree	None
3	agromics	Jamboree: summary of article	1 File
Add Delete Open			

Figure 12. Grille des commentaires de l'interface *Projet*.

Ce formulaire permet de modifier les informations du (nouveau) commentaire : le titre et le texte. L'utilisateur du commentaire est nécessairement celui qui est connecté, celui qui crée donc le commentaire.

Add new comment

Title:

User :

agromics

Text:

Attachments:

Add

Details

Delete

Save

Cancel

Figure 13. Formulaire d'ajout d'un nouveau commentaire.

Il est également possible de créer de nouvelles pièces jointes au commentaire depuis le panneau de gauche (« Attachments »), via le bouton « Add ». Ceci provoque l'ouverture d'un nouveau formulaire, dans une nouvelle fenêtre (**Figure 14**). Via ce formulaire, il est possible d'entrer les informations de la pièce jointe : son nom, son auteur (facultatif) et sa description (facultatif). Ensuite, soit on sélectionne une pièce jointe à télécharger sur le serveur (cas 1), soit on entre une URL d'un fichier déjà en ligne (cas 2). Lorsqu'on clique sur « Save », la pièce jointe est ajoutée au commentaire côté client. Dans le cas 1 mentionné précédemment, le fichier est téléchargé dans le dossier tempo-

raire du serveur. Lorsqu'on clique sur « *Save* » du formulaire du commentaire, le commentaire est sauvegardé sur le serveur, de même que toutes ses pièces jointes, et les éventuels fichiers (cas 1) sont transférés du dossier temporaire au dossier utilisateur sur le serveur. Une fois ajoutée, les détails de la pièce jointe peuvent être modifiés en la sélectionnant puis en cliquant sur le bouton « *Details* » (**Figure 13**). Double-cliquer sur une pièce jointe ouvre l'URL associée. Ceci peut également se faire via le bouton « *Open* ».

Figure 14. Formulaire d'ajout d'une pièce jointe à un commentaire.

C. FICHE VOIE METABOLIQUE

Objectifs

Pour corriger manuellement un réseau métabolique, il est important d'avoir une information la plus complète possible sur chaque objet de ce réseau. Cette fiche a pour objectif de mettre en évidence au sein d'une même interface ce qui relie une voie métabolique aux autres objets du réseau : les réactions qu'elle contient et les gènes qu'elle implique. L'utilisateur a la possibilité de voter pour chaque voie métabolique via cette interface pour indiquer s'il pense qu'elle existe et qu'elle est correcte ou non. Il peut ajouter des commentaires pour donner plus de détails sur son avis.

Fonctionnement

La fiche s'ouvre lorsque l'on clique sur l'icône  d'une voie métabolique dans la grille *Pathways* de l'onglet *Network Data* (**Figure S10** en annexe). La fiche s'organise en panneaux superposés en accordéon (**Figure 15**). Le premier panneau affiche les réactions incluses dans la voie métabolique. Le second panneau est la liste des gènes qui y sont également inclus. Ensuite, sont accessibles les

commentaires de la réaction. Il s'agit de la même interface que pour les commentaires d'un *Projet* (voir paragraphe B-2). Le dernier panneau permet de voter pour la voie métabolique (voir paragraphe E pour sa description).

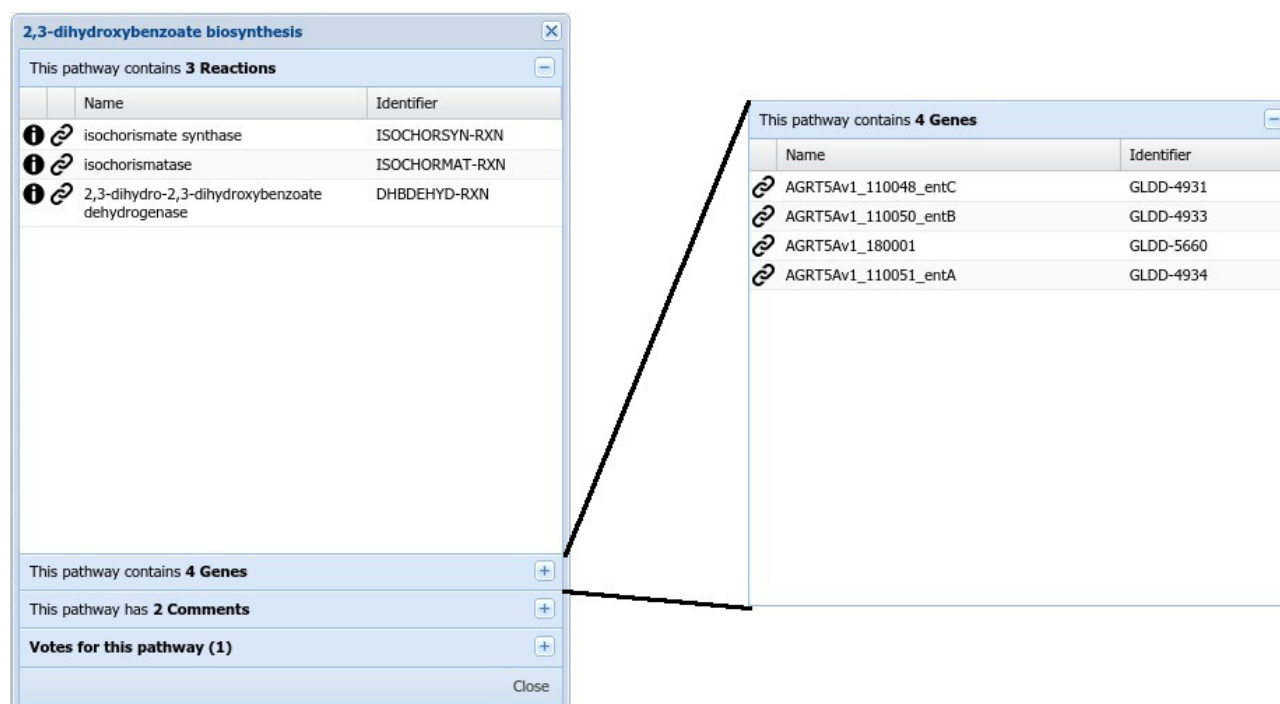


Figure 15. Fiche d'une Voie Métabolique. À droite, le panneau des gènes contenus dans la réaction

D. FICHE REACTION

Objectifs

De la même manière que pour la fiche Voie Métabolique, cette fiche a pour objectif de mettre en évidence au sein d'une même interface ce qui relie une réaction aux autres objets du réseau : la formule de la réaction permet de la relier aux métabolites qu'elle consomme ou produit, et deux grilles permettent de voir les voies métaboliques dans laquelle elle est retrouvée et quels gènes sont impliqués dans cette réaction. L'utilisateur a là aussi la possibilité de voter pour chaque réaction via cette interface pour indiquer s'il pense qu'elle existe et qu'elle est correcte ou non. Il peut ajouter des commentaires pour donner plus de détails sur son avis.

Fonctionnement

La fiche s'ouvre lorsque l'on clique sur l'icône **i** d'une réaction dans la grille *Réactions* de l'onglet *Network Data* (**Figure S10** en annexe). Cette fiche (**Figure 16**) est construite de la même manière que la fiche Voie Métabolique. Le premier panneau, ouvert par défaut, permet de visualiser l'équation de la réaction. Cette équation apparaît avec le nom des métabolites, leur identifiant ou leur formule chimique. Certains métabolites n'ont pas de formule chimique définie dans la base de données. Dans ce cas, on la remplace par *NA*. Dans le panneau en dessous, on peut visualiser

l'association GPR (voir paragraphe F ci-dessous pour plus de détails). Les panneaux du dessous affichent les voies métaboliques dont fait partie la réaction et les gènes qui codent pour les protéines formant les complexes enzymatiques qui catalysent la réaction. Enfin, il y a les mêmes panneaux des commentaires et des votes que dans la fiche Voie Métabolique.

2-oxoglutarate dehydrogenase complex

Equations of the reaction

Equation with names:
 $1 \text{ coenzyme A} + 1 \text{ NAD}^+ + 1 \text{ 2-oxoglutarate} \rightarrow 1 \text{ CO}_2 + 1 \text{ succinyl-CoA} + 1 \text{ NADH}$

Equation with identifiers:
 $1 \text{ CO-A} + 1 \text{ NAD} + 1 \text{ 2-KETOGLUTARATE} \rightarrow 1 \text{ CARBON-DIOXIDE} + 1 \text{ SUC-COA} + 1 \text{ NADH}$

Equation with chemical formulas:
 $1 \text{ C}_{21}\text{H}_{32}\text{N}_7\text{O}_{16}\text{P}_3\text{S}_1 + 1 \text{ C}_{21}\text{H}_{26}\text{N}_7\text{O}_{14}\text{P}_2 + 1 \text{ C}_5\text{H}_4\text{O}_5 \rightarrow 1 \text{ C}_{10}\text{O}_2 + 1 \text{ C}_{25}\text{H}_{35}\text{N}_7\text{O}_{19}\text{P}_3\text{S}_1 + 1 \text{ C}_{21}\text{H}_{27}\text{N}_7\text{O}_{14}\text{P}_2$

GPR association viz

This reaction exists in 2 Pathways

Name	Identifier
TCA cycle I (prokaryotic)	TCA
TCA cycle II (eukaryotic)	PWY-5690

This reaction involves 4 Genes

Name	Identifier
AGRT5Av1_70265_sucA	GLDD-3920
AGRT5Av1_70269_lpd	GLDD-3924
AGRT5Av1_70266_sucB	GLDD-3921
AGRT5Av1_20698_lpd	GLDD-1434

Figure 16. Fiche d'une Réaction. En haut à droite de la figure : panneau des voies métaboliques qui incluent la réaction. En bas à droite de la figure : panneau des gènes qui codent pour les enzymes qui catalysent la réaction.

E. PANNEAU VOTES

Objectifs

L'objectif est de proposer un outil, à ceux qui n'oseraient pas modifier la base de données, ou qui n'en ont pas le droit, permettant de donner leur avis sur un objet du réseau métabolique. Également, l'idée est de pouvoir rassembler facilement les avis de plusieurs personnes pour chaque objet du réseau métabolique, pour faciliter l'organisation de la correction manuelle du réseau métabolique.

Fonctionnement

Le panneau Votes (**Figure 17**) est inclus dans toutes les fiches des objets d'un réseau métabolique. Un code couleur a été décidé pour l'ensemble des vues représentant les votes. Ainsi, le vert correspond au vote « *L'objet existe dans cet organisme et est correct* ». Le jaune correspond au vote « *L'objet existe dans cet organisme mais comporte des erreurs* ». Enfin, le rouge correspond au vote « *L'objet n'existe pas dans cet organisme* ». Le premier onglet de ce panneau permet à l'utilisateur de voter pour l'objet, s'il en a le droit. Sinon ce premier onglet n'est pas affiché. Le choix de l'utilisateur est en gras souligné, en noir une fois le choix enregistré dans la base de don-

nées du serveur. Lors du clic, le choix est en gris le temps que l'enregistrement se fasse sur le serveur. Une fois que ce dernier renvoie une réponse positive, le choix passe en noir.

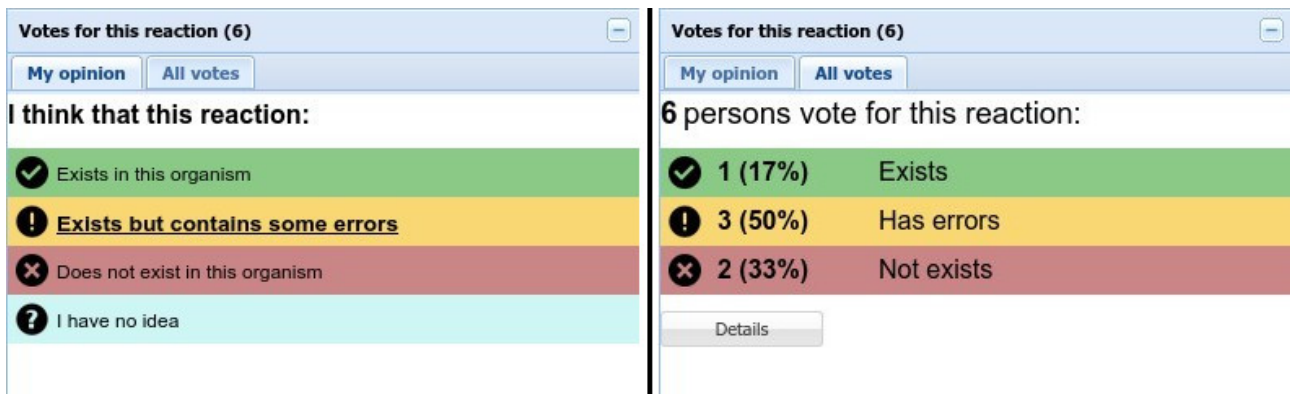


Figure 17. Panneau Votes d'une Réaction. Les panneaux Votes des autres objets sont identiques, seul le mot « reaction » est changé pour le nom de l'objet concerné dans le panneau.

Le second onglet récapitule les votes de tous les utilisateurs (**Figure 18**). Le bouton « Details » permet de savoir qui a voté pour quoi. Dans les grilles de l'onglet *Network Data*, le bouton « Curation Votes » permet d'ajouter une colonne, « Votes summary », qui résume les votes de tous les objets de la grille, en permettant de les classer de ceux qui ont le plus de votes à ceux qui en ont le moins (**Figure 19**).

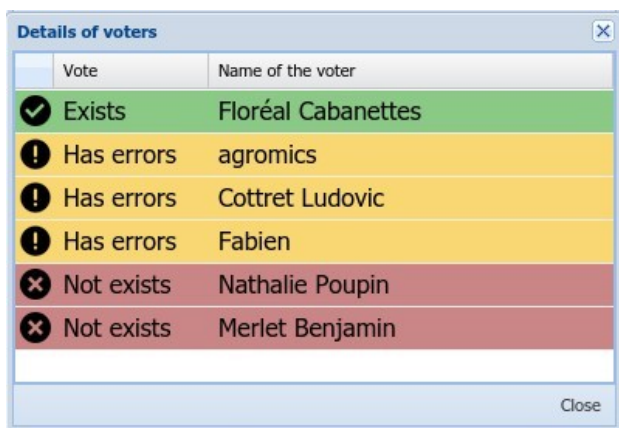


Figure 18. Détails des votants d'une réaction.

Cette colonne représente les votes résumés pour chaque réaction, sous la forme d'une barre multicolore. Cette barre reprend les couleurs des autres vues, et la surface (soit la largeur) de chaque rectangle de la barre est proportionnel au nombre de votants correspondant, indiqué en son centre. Si l'utilisateur clique sur l'un de ces graphiques, la fiche de l'objet correspondant (la réaction correspondante dans le cas de la **Figure 19**) s'ouvre avec le panneau Votes ouvert.

← BioSources Compartments (1/1) Pathways (377/377) Reactions (1740/1740)					
+ Add Edit ✕ Delete ☁ Commit Changes ☰ Multiple affectation 🗳 Curation 🗳 Votes					
		Name	Identifier	E.C.	Votes summary ▾
1	🔍	2-oxoglutarate dehydrogenase...	2OXOGLUTARATE...	NA	1 3 2
2	🔍	α-Arabinofuranosidase	3.2.1.55-RXN	3.2.1.55	1
3	🔍	α-galactosidase	ALPHAGALACTOSI...	3.2.1.22	1
4	🔍	α-amylase	ALPHA-AMYL-RXN	3.2.1.1	1
5	🔍	DEOXYADENPHOSPHOR-RXN	DEOXYADENPHOS...	2.4.2.1	No votes
6	🔍	β-Acetylhexosaminidase	3.2.1.52-RXN	3.2.1.52	No votes

Figure 19. Ajout de la colonne résumant les votes des utilisateurs. La première ligne correspond à la même réaction que celle de la **Figure 18**.

F. VISUALISATION DE L'ASSOCIATION GENE-PROTEINE-REACTION (GPR) DANS LA FICHE REACTION

Objectifs

La grille des gènes, incluse dans la fiche Réaction, permet de voir les gènes impliqués dans une réaction. Cependant, elle ne permet pas de les relier aux protéines, et de savoir quelles protéines sont essentielles pour que la réaction se fasse. En effet, si deux protéines, donc deux gènes, sont nécessaires, soit ces 2 protéines forment deux enzymes distinctes capables de catalyser la réaction indépendamment l'une de l'autre, soit elles forment un complexe enzymatique qui catalyse la réaction et dans ce cas les deux protéines sont nécessaires pour que la réaction soit possible. Grâce à la visualisation GPR sous forme de graphe, ceci est facilement identifiable. L'objectif est également de savoir quelles autres réactions sont catalysées par les enzymes qui catalysent la réaction d'intérêt. Ceci dans un but de corrections manuelles d'un réseau métabolique, afin d'identifier les réactions potentiellement fausses dans le réseau (voir exemple ci-dessous).

Fonctionnement

La visualisation est générée à l'ouverture de la fiche *Réaction*. Elle représente l'association entre les gènes, les protéines et les réactions, sous forme de graphe (**Figure 20**). Ce graphe représente les réactions sous forme de rectangles, foncé pour la réaction à partir de laquelle est demandée la visualisation, clairs pour les autres réactions catalysées par les enzymes qui catalysent la première. Les complexes enzymatiques sont des doubles losanges violets, mais ne s'affichent que s'ils sont formés d'au moins 2 protéines, représentées par des losanges bleus. Dans le cas contraire, la protéine est également l'enzyme et est reliée directement aux réactions. Enfin, les gènes sont représentés sous forme d'hexagones verts.

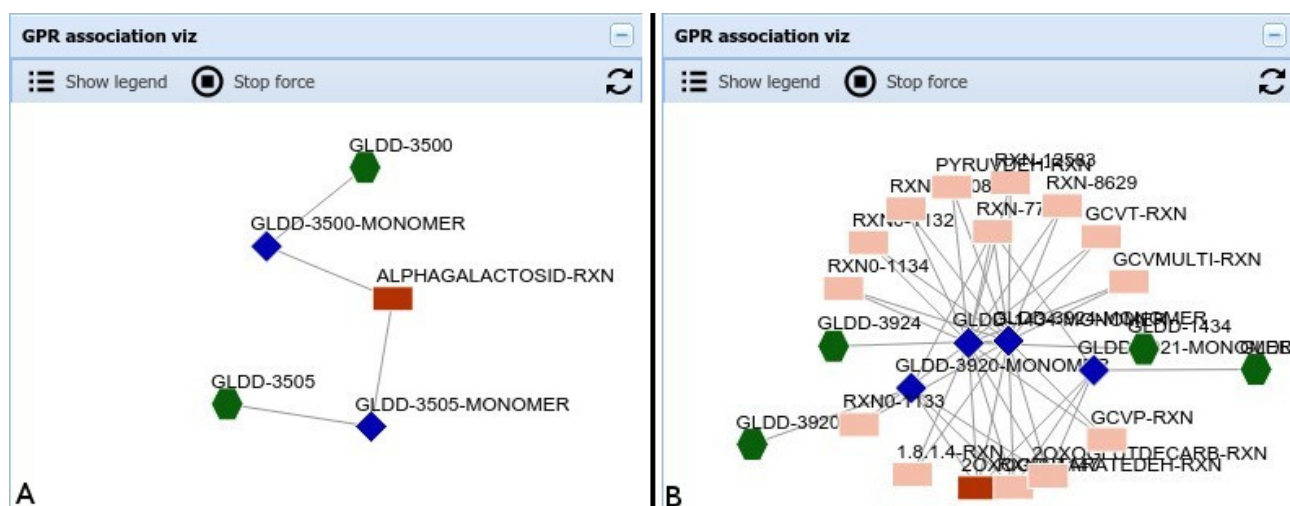


Figure 20. Visualisation GPR réalisées à partir de 2 réactions. Les rectangles correspondent aux réactions. En foncé la réaction à partir de laquelle la visualisation GPR est effectuée, en clair les autres réactions catalysées par les enzymes qui catalysent la première réaction. En bleu les protéines qui sont dans ce cas également les enzymes, et en vert les gènes.

Cette visualisation peut permettre d'identifier facilement certaines erreurs présentes dans le réseau. Ainsi, du fait de la spécificité d'une enzyme, on peut s'attendre à ce qu'elle catalyse un nombre restreint de réactions, et qu'une réaction soit catalysée également par un nombre restreint d'enzymes (**Figure 20A**). Cependant, pour certaines réactions, la visualisation GPR montre qu'un groupe d'enzymes catalysent un grand nombre de réactions (**Figure 20B**). Ceci pointe donc vers des erreurs probables dans le réseau et permet d'orienter la correction manuelle du réseau.

G. « COMPLETUDE » DES VOIES METABOLIQUES

Objectifs

Du fait de la difficulté d'inférer automatiquement les voies métaboliques sur un réseau, les reconstructions métaboliques automatiques contiennent un certain nombre de faux positifs au niveau des voies métaboliques, c'est-à-dire des voies métaboliques qui sont incluses dans le réseau reconstruit mais qui ne sont pas présentes réellement chez l'organisme. Il est possible d'en éliminer une partie en regardant pour chaque voie métabolique la proportion de réactions n'étant pas catalysées par une enzyme par rapport au nombre de réactions totales qu'elle contient. C'est ce que l'on appelle la *complétude* des voies métaboliques. Plus elle est faible, plus on peut penser que la voie métabolique correspondante est erronée. À l'inverse, si elle est inférieure à 100% mais proche, il est probable qu'il manque certaines enzymes dans le réseau reconstruit. Ceci donne donc une première idée des voies métaboliques à vérifier.

Fonctionnement

La complétude est présente dans la grille *Pathways* de l'onglet *Network Data* de *MetExplore* (**Figure S10** en annexe). Elle apparaît dans une colonne masquée par défaut, mais peut être affichée grâce au menu de l'en-tête de la grille, en cochant la colonne « % Reactions with Enz » (**Figure 21**).

← BioSources Compartments (1/1) Pathways (377/377) Reactions (1740/1740) Metabolites (18) →				
+ Add Edit ✕ Delete 📊 Curation Statistics 🗳 Curation Votes				
		Name ^	Identifier	% Reactions with Enz
39	 	adenosylcobalamin biosynthesis from cobyri...	PWY-5509	62 %
40	 	adenosylcobalamin biosynthesis from cobyri...	PWY-5508	55 %
41	 	adenosylcobalamin biosynthesis I (early cobalt...	PWY-5507	52 %
42	 	adenosylcobalamin biosynthesis II (late cobalt...	P381-PWY	76 %
43	 	adenosylcobalamin salvage from cobalamin	PWY-6268	100 %
44	 	adenosylcobalamin salvage from cobinamide I	COBALSYN-PWY	83 %
45	 	aerobic respiration (cytochrome c)	PWY-3781	100 %
46	 	alanine biosynthesis I	ALANINE-VALINESYN-PWY	66 %
47	 	alanine biosynthesis III	PWY0-1021	100 %
48	 	alanine degradation I	ALADEG-PWY	100 %
49	 	allantoin degradation to glyoxylate III	PWY-5705	50 %
50	 	allantoin degradation to ureidoglycolate II (a...	PWY-5698	33 %
51	 	ammonia assimilation cycle I	PWY-6963	100 %
52	 	ammonia assimilation cycle II	PWY-6964	100 %
53	 	ammonia oxidation I (aerobic)	AMMOXID-PWY	50 %
54	 	androstenedione degradation	PWY-6944	6 %
55	 	arginine biosynthesis I	ARGSYN-PWY	100 %
56	 	arginine biosynthesis II (acetyl cycle)	ARGSYNBSUB-PWY	100 %
57	 	arginine degradation I (arginase pathway)	ARGASEDEG-PWY	50 %
58	 	arginine degradation II (AST pathway)	AST-PWY	20 %
59	 	arginine degradation III (arginine decarboxyla...	PWY0-823	50 %

Figure 21. Affichage de la complétude des voies métaboliques dans la grille *Pathways* de l'onglet *Network Data*. La Complétude s'affiche dans la colonne nommée « % Reactions with Enz », masquée par défaut. Dans cette figure, on voit une partie des voies métaboliques de la reconstruction BioCyc d'*Agrobacterium tumefaciens* C58 du 8 juillet 2014.

Dans cette colonne, les *complétudes* inférieures à 25% sont affichées en rouge, celles comprises entre 25% et 49% sont affichées en orange, celles comprises entre 50% et 74% sont affichées en vert clair. Enfin, celles supérieures ou égales à 75% sont affichées en vert vif. Il peut être particulièrement intéressant ici d'accéder à la fiche des voies métaboliques à *complétude* faible, grâce au bouton de la première colonne, pour visualiser les détails de cette voie métabolique, et notamment les réactions qu'elle contient. D'autant que depuis cette grille des réactions, on peut accéder à la fiche Réaction de chacune (**Figure 15**).

Un résumé des *complétudes* (**Figure 22**) est également disponible en cliquant sur le bouton « *Curation statistics* » dans la barre d'outils du haut de la grille *Pathways* de l'onglet *Network Data* de *MetExplore* (**Figure S10** en annexe). Ce résumé se présente sous la forme d'un histogramme représentant pour chaque catégorie mentionnée ci-dessus (<25%, ≥ 25%, ≥50% et ≥75%), le nombre de voies métaboliques de la *BioSource*. On peut également afficher ces valeurs sous forme de pourcentage dans le panneau du bas (fermé par défaut). Dans la barre d'outils du bas, le menu *Export* permet d'exporter le graphique courant (avec les valeurs absolues, par défaut, ou les pourcentages) en vectoriel (svg) ou png.

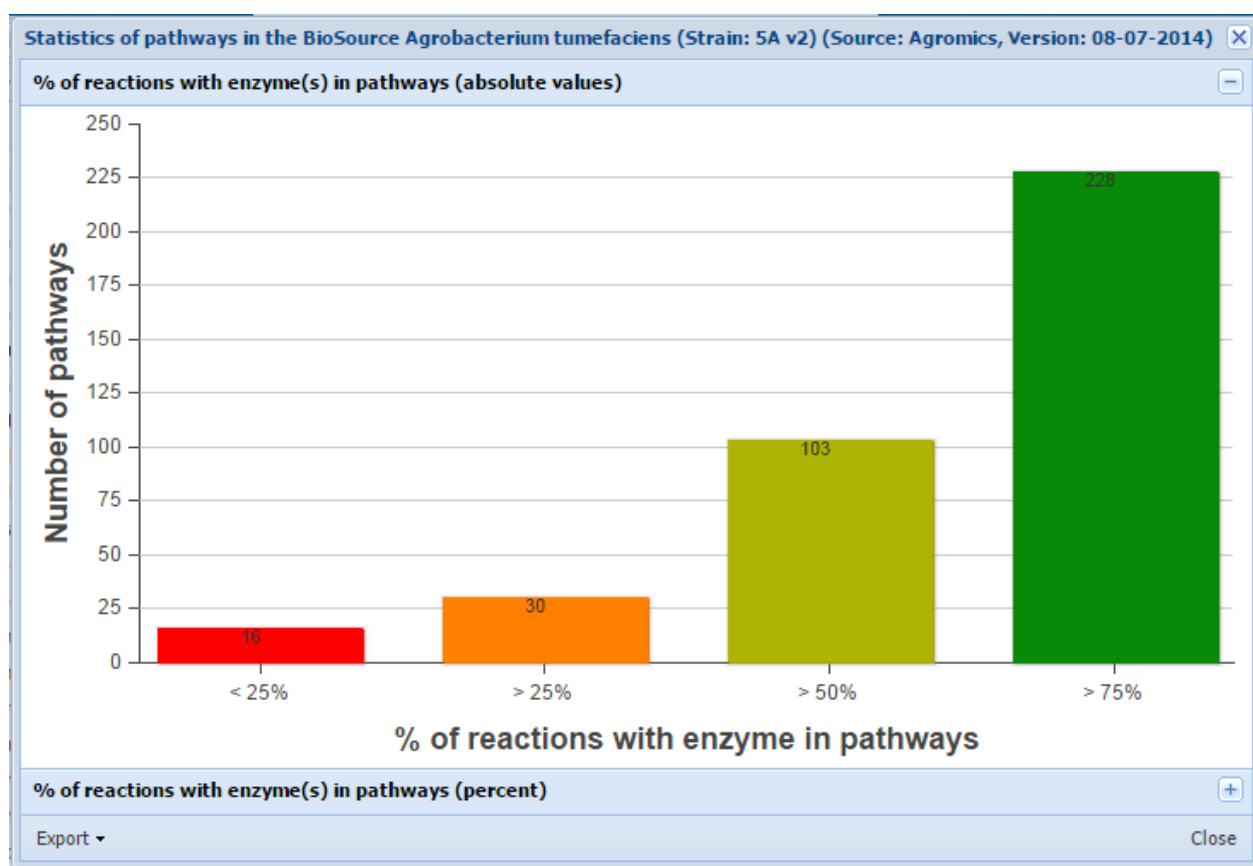


Figure 22. Résumé de la *complétude* des voies métaboliques d'une *BioSource*.

H. EXPORT DE TOUT LE RESEAU AU FORMAT EXCEL

Objectifs

Nous avons pour objectif de développer des outils simples dans *MetExplore*, pour qu'un plus grand nombre d'utilisateurs puissent les utiliser sans besoin de connaissances particulières en informatique. Cependant, comme dit précédemment dans le rapport, tous les participants au projet *Agromics* n'auront pas les mêmes facilités avec l'interface de *MetExplore*. À l'inverse, une feuille Excel est quelque chose que tout le monde connaît et sait – un minimum – manipuler. C'est pourquoi nous proposons l'export de tout le réseau au format Excel, comme solution de secours. Également, cette fonctionnalité permet de travailler sur le réseau sans connexion internet.

Fonctionnement

L'export se fait en un clic grâce à un menu de *MetExplore*. Quand l'export est demandé, les données de toutes les grilles présentes dans *Network Data* (**Figure S10** en annexe) sont collectées, en conservant l'ordre des grilles et des colonnes qui les composent, et sont envoyées au serveur qui s'occupe de générer le fichier Excel qui est ensuite proposé en téléchargement à l'utilisateur. Ce fichier contient alors toutes les grilles dans une feuille distincte d'Excel. La largeur des colonnes s'adapte au mieux à leur contenu. Le nom de colonne est placé dans un en-tête, en gras.

BILAN ET PERSPECTIVES

Lors de ce stage, j'ai donc amélioré *MetExplore Annotation* en lui ajoutant :

1. Une interface collaborative grâce aux outils utilisateur et à la gestion de *Projets*
2. Un système de gestion des droits afin de les adapter au rôle de chaque collaborateur d'un *Projet*.
3. Différents outils d'évaluation du réseau métabolique afin de faciliter son affinement manuel.
4. Un système permettant à tous les utilisateurs de donner leur opinion sur chaque objet qui compose le réseau métabolique.
5. Favoriser la participation de tous les utilisateurs à l'annotation métabolique.

La reconstruction métabolique automatique, bien que rapide, est montrée par les spécialistes comme contenant beaucoup d'erreurs (Thiele & Palsson, 2010a; Swainston *et al*, 2011). Ces reconstructions ne sont pas suffisamment bonnes pour être publiables. Il est nécessaire de faire des corrections manuelles sur ces réseaux, car toutes les étapes ne sont pas automatisables (Thiele & Palsson, 2010a).

Cependant, ces réseaux reconstruits automatiquement sont précieux. Ils constituent une version *brouillon*, qui peut servir de point de départ. Dans le cadre du projet *agromics*, qui a pour vocation la reconstruction métabolique de plusieurs espèces d'*Agrobacterium* vivant dans des environnements très différents, on souhaite utiliser ces reconstructions automatiques et les affiner manuellement pour corriger au mieux leurs erreurs.

Aussi, le challenge des bioinformaticiens est de proposer des outils qui permettent de faciliter l'annotation et la correction manuelle d'un réseau métabolique. C'est dans cette optique que s'inscrit *MetExplore Annotation*, dans lequel mon stage s'insère. Nous utilisons la capacité qu'a *MetExplore* d'importer directement un réseau KEGG depuis leur plateforme, ou un réseau BioCyc au travers d'un fichier SBML. Nous proposons ensuite une interface permettant d'apporter manuellement des corrections à ce réseau automatique, le plus intuitivement possible, car soucieux d'une

bonne interaction homme-machine (IHM). Lors de mon stage, j'ai répondu à différents objectifs, détaillés dans la suite de la discussion.

1. FAVORISER L'ANNOTATION COLLABORATIVE

Afin de favoriser l'annotation **collaborative**, j'ai mis en place dans le cadre de mon stage un système de *Projets* et une véritable interface utilisateur, une originalité de *MetExplore Annotation* parmi les outils d'annotation métabolique (**Tableau 1**). En regroupant différentes vues du travail fait et à effectuer, au travers de la *TODO List* et de l'historique notamment, chaque utilisateur peut donc suivre facilement le travail effectué par les différents membres du *Projet* auquel il participe. Et si ce système sert aux participants actifs du *Projet*, les organisateurs qui seraient impliqués dans de nombreux *Projets* et y ayant essentiellement un rôle d'arbitre peuvent voir en un coup d'œil les avancées et diriger les utilisateurs en conséquence, facilement.

2. DROITS UTILISATEURS

La création de droits d'accès aux réseaux métaboliques différents pour chaque utilisateur permet d'adapter l'accès de chaque utilisateur au rôle qu'il apporte dans la reconstruction de chaque réseau métabolique. Ainsi, les spécialistes peuvent donner leur opinion sur le réseau tandis que les organisateurs font la synthèse de ces opinions et modifient le réseau en conséquence.

3. CREATION D'OUTILS D'EVALUATION DU RESEAU METABOLIQUE

Par ailleurs, j'ai mis en place différents outils permettant de mettre en évidence les anomalies présentes dans un réseau métabolique. À commencer par la visualisation des relations Gène-Protéine-Réaction (GPR). Actuellement, la visualisation GPR n'est accessible qu'à partir d'une réaction donnée. Il serait intéressant par la suite de la rendre possible à partir des complexes enzymatiques, des protéines ou des gènes. Ceci ne demande que peu de modifications du code existant et serait réalisable à court terme. Cette visualisation est déjà présente sur des sites comme BioCyc, qui nous a inspiré l'idée. Cependant, nous avons développé une visualisation plus interactive, qui permet de repositionner les éléments sur le dessin, quand les associations sont complexes. Nous avons également créé des outils permettant d'évaluer les voies métaboliques, au travers de la *complétude* de ces voies. Ceci existe déjà dans *Pathway tools* (Karp *et al*, 2002), et est un outil important du fait du nombre de voies métaboliques ajoutées par erreur dans une reconstruction automatique.

Ces outils seront un plus pour corriger une reconstruction. Les collaborateurs du *Projet* pourront également utiliser d'autres outils déjà présents dans *MetExplore*, comme la visualisation du réseau métabolique sous forme de graphe, ou encore les différentes analyses de flux, pour mettre en lumière d'autres anomalies. Mais d'autres outils seraient et seront les bienvenus dans *MetExplore Annotation*, dans ce but. Notamment, des outils pour détecter et corriger des trous (« *gaps* ») dans le

réseau. De tels outils existent, comme GapFind, capable de détecter les trous par détection de réactions bloquées (analyses de flux) et GapFill capable de les remplir (Kumar *et al*, 2007), en modifiant le réseau (rend des réactions réversibles, ...) ou en allant chercher dans d'autres réseaux proches dans la base de données, ou encore en ajoutant des réactions de transport. SMILEY (Reed *et al*, 2006) utilise lui les données de croissance cellulaire pour détecter et remplir les *gaps*. Nous pourrions nous inspirer de ces outils afin de les intégrer à *MetExplore*. Également, des outils ont déjà été développés par l'équipe Bioinformatique du LIPM dans laquelle j'effectue ce stage. Ils sont en ligne de commande, concernent le *gap filling* et d'autres analyses aidant la correction manuelle d'un réseau métabolique. Ils pourraient être intégrés à *MetExplore* dans le futur grâce à la création d'une interface graphique dédiée.

4. FAVORISER LE PARTAGE D'OPINIONS

Il est primordial de noter qu'une annotation de qualité nécessite de réunir des spécialistes de divers horizons, ayants chacun leurs connaissance précise mais limitée à un certain domaine. Mais en les réunissant, leurs domaines se complètent et permettent de recouvrir un maximum des connaissances nécessaires à l'élaboration d'une reconstruction métabolique de qualité. Aussi, il peut s'avérer utile de confronter les avis sur un composant du réseau avant que les organisateurs fassent finalement les modifications décidées. Afin de faciliter cette confrontation, j'ai donc mis en place lors de mon stage un système de votes et de commentaires. Grâce aux codes graphiques utilisés, le système de votes est intuitif, et la colonne résumé des votes permet aux organisateurs de voir rapidement une vue d'ensemble des objets ayant reçu un ou plusieurs votes. Ce système de vote s'inspire de sites comme *Stack Overflow*¹, et constitue une originalité de *MetExplore Annotation* (**Tableau 1**).

Au niveau du système de commentaires utilisé pour les *Projets* et les différents objets du réseau métabolique, quelques améliorations pourraient exister. Notamment, il serait nécessaire de pouvoir répondre à un commentaire, et que la réponse s'affiche en tant que telle dans l'interface. Aussi, il serait intéressant de pouvoir associer une catégorie aux commentaires, avec un code couleur dans la liste des commentaires. Ces modifications ne demandent que peu de changements au niveau du code et pourraient être mises en place rapidement.

5. RENFORCER L'ANNOTATION POUR TOUS

Mais en plus de ces outils, il a fallu prévoir que certains seraient malgré tout plus à l'aise avec un outil qu'ils connaissent déjà bien, comme un fichier Excel. Également, il fallait rendre possible la modification hors connexion d'un réseau métabolique. C'est pourquoi j'ai implémenté lors de mon stage l'export de tout un réseau dans un fichier Excel. Cependant, on perd lors de cet export le lien

¹ <http://stackoverflow.com>

entre la réaction et les métabolites. Ceci pourrait être corrigé en exportant la formule de la réaction dans une colonne de la feuille des réactions. Les relations GPR sont également perdues. Elles pourraient également être codées dans une nouvelle colonne des feuilles Excel. Aussi, une fois les modifications apportées sur le fichier Excel, elles doivent être actuellement reportées manuellement sur *MetExplore* par les organisateurs du *Projet*. Cependant, l'import automatique est en cours d'implémentation par l'équipe de *MetExplore* et devrait être disponible rapidement.

CONCLUSION ET PERSPECTIVES

Tout le développement logiciel que j'ai réalisé lors de ce stage sera mis à profit lors du *jamboree* qui sera organisé par mon encadrant et moi-même les 16 et 17 juin prochain. Ils permettront de faciliter la correction des trois reconstructions automatiques du réseau d'*Agrobacterium tumefaciens* C58 retenues, par un ensemble de spécialistes de différentes disciplines. Les outils graphiques développés, tels que la colonne résumant les votes, les graphiques résumant la *complétude* des voies métaboliques, ou encore l'historique permettra à la fin de chaque journée de faire le point sur le travail accompli.

Ainsi, ce stage m'a permis de développer dans un langage que j'avais besoin d'approfondir, le *JavaScript*, et en utilisant l'environnement *ExtJS*. L'apprentissage de cet environnement a pris environ 1 mois de mon stage avant de le maîtriser. Aussi, le travail en équipe, et les nombreux échanges m'ont aidé à progresser dans le domaine de la bioinformatique. Tout le travail réalisé a été fait dans l'optique de la reconstruction d'espèces d'*Agrobacterium* et du *jamboree*, mais pourra être utilisée dans d'autres *Projets* par d'autres utilisateurs. Aussi, je présenterai *MetExplore Annotation* aux Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM) à Clermont-Ferrand, qui ont lieu du 6 au 9 juillet prochain.

MetExplore contient donc la plupart des outils présents dans les logiciels actuels permettant l'affinement manuel de réseaux métaboliques, et les complète de plusieurs outils innovants (**Tableau 1**). La suite du développement de *MetExplore Annotation* consistera en l'ajout de nouvelles fonctionnalités ayant toujours pour objectif de faciliter l'affinement manuel d'un réseau métabolique. Une fonctionnalité particulièrement attendue sera de pouvoir comparer plusieurs réseaux métaboliques, fonctionnalité qui est rarement présente dans les outils actuels (**Tableau 1**). Ainsi, *MetExplore Annotation* devrait permettre d'obtenir des réseaux métaboliques de plus en plus précis, contenant de moins en moins d'erreurs.

BIBLIOGRAPHIE

- Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I & Nielsen J (2013) The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for *Penicillium chrysogenum*. *PLoS Comput. Biol.* **9**: e1002980
- Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ & Herrgard MJ (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* **2**: 727–738
- Bostock M, Ogievetsky V & Heer J (2011) D³ Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.* **17**: 2301–2309
- Cottret L, Wildridge D, Vinson F, Barrett MP, Charles H, Sagot M-F & Jourdan F (2010) MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.* **38**: W132–W137
- Henry CS, DeJongh M, Best AA, Frybarger PM, Lindsay B & Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* **28**: 977–982
- Karp PD, Paley S & Romero P (2002) The Pathway Tools software. *Bioinformatics* **18**: S225–S232
- Kumar VS, Dasika MS & Maranas CD (2007) Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* **8**: 212
- Liao Y-C, Tsai M-H, Chen F-C & Hsiung CA (2012) GEMSiRV: a software platform for GENome-scale metabolic model simulation, reconstruction and visualization. *Bioinformatics* **28**: 1752–1758
- Liberal R, Lisowska BK, Leak DJ & Pinney JW (2015) PathwayBooster: a tool to support the curation of metabolic pathways. *BMC Bioinformatics* **16**: Available at: <http://www.biomedcentral.com/1471-2105/16/86> [Accessed May 19, 2015]
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC & Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**: W182–W185
- Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS & Palsson BO (2006) Systems approach to refining genome annotation. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 17480–17484
- Swainston N, Smallbone K, Mendes P, Kell D & Paton N (2011) The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. *J Integr Bioinform* **8**: 186
- Thiele I & Palsson BØ (2010a) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**: 93–121
- Thiele I & Palsson BØ (2010b) Reconstruction annotation jamborees: a community approach to systems biology. *Mol. Syst. Biol.* **6**: Available at: <http://msb.embopress.org/cgi/doi/10.1038/msb.2010.15> [Accessed May 17, 2015]
- Thorleifsson SG & Thiele I (2011) rBioNet: A COBRA toolbox extension for reconstructing high-quality biochemical networks. *Bioinformatics* **27**: 2009–2010
- ExtJS (version 4.2.1). JavaScript. Sencha, 2013. <http://www.sencha.com/products/extjs/>.

ANNEXES

FIGURES SUPPLEMENTAIRES

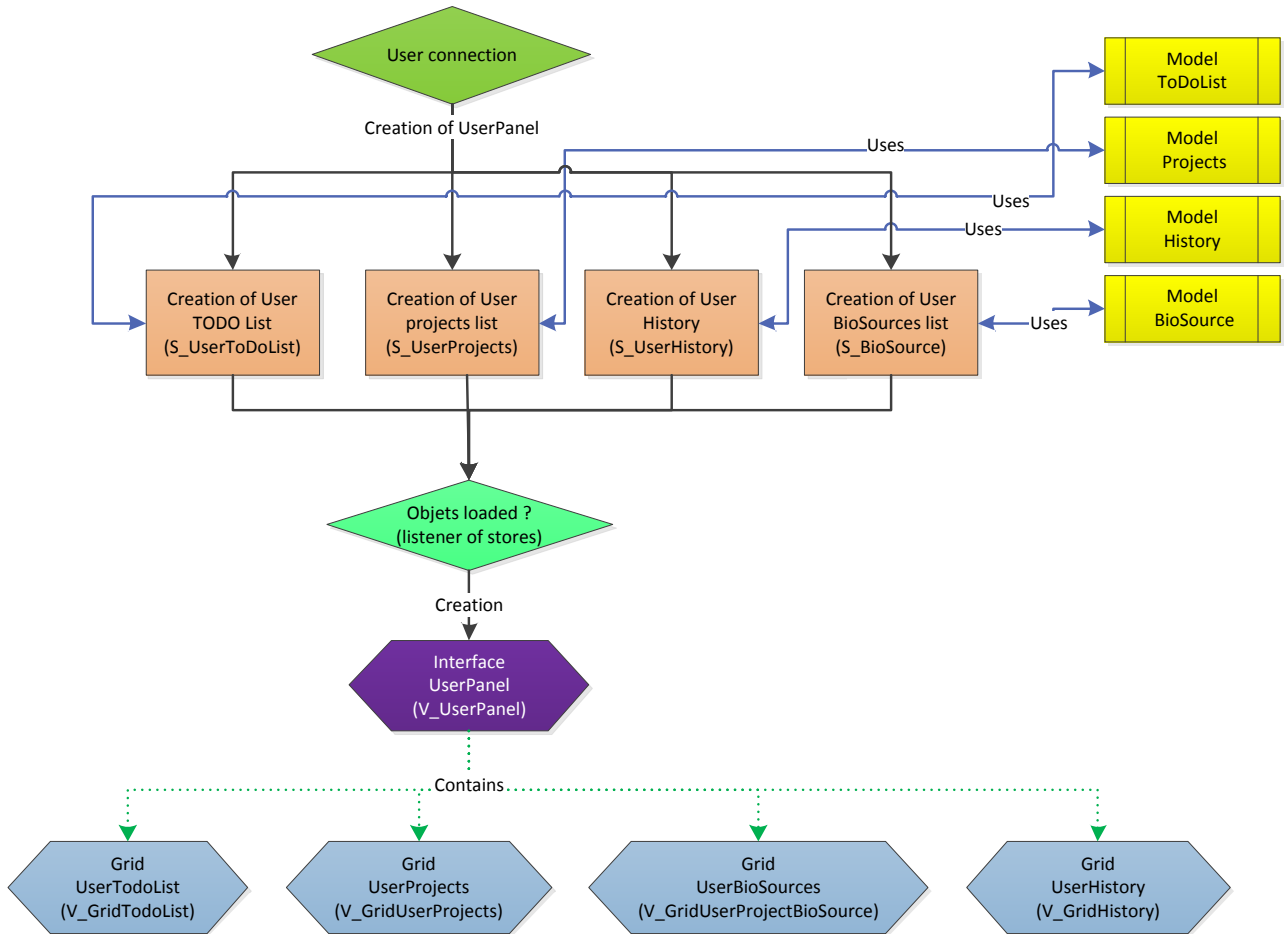


Figure S1. Connection d'un utilisateur à *MetExplore* : initialisation du panneau *Utilisateur*. Entre parenthèse est indiquée la Classe ExtJS concernée. En violet les super-vues contenant d'autres vues, en bleu les vues, en vert les contrôleurs et en marron les stores. Les Stores sont remplis depuis la base de donnée du serveur via des scripts PHP.

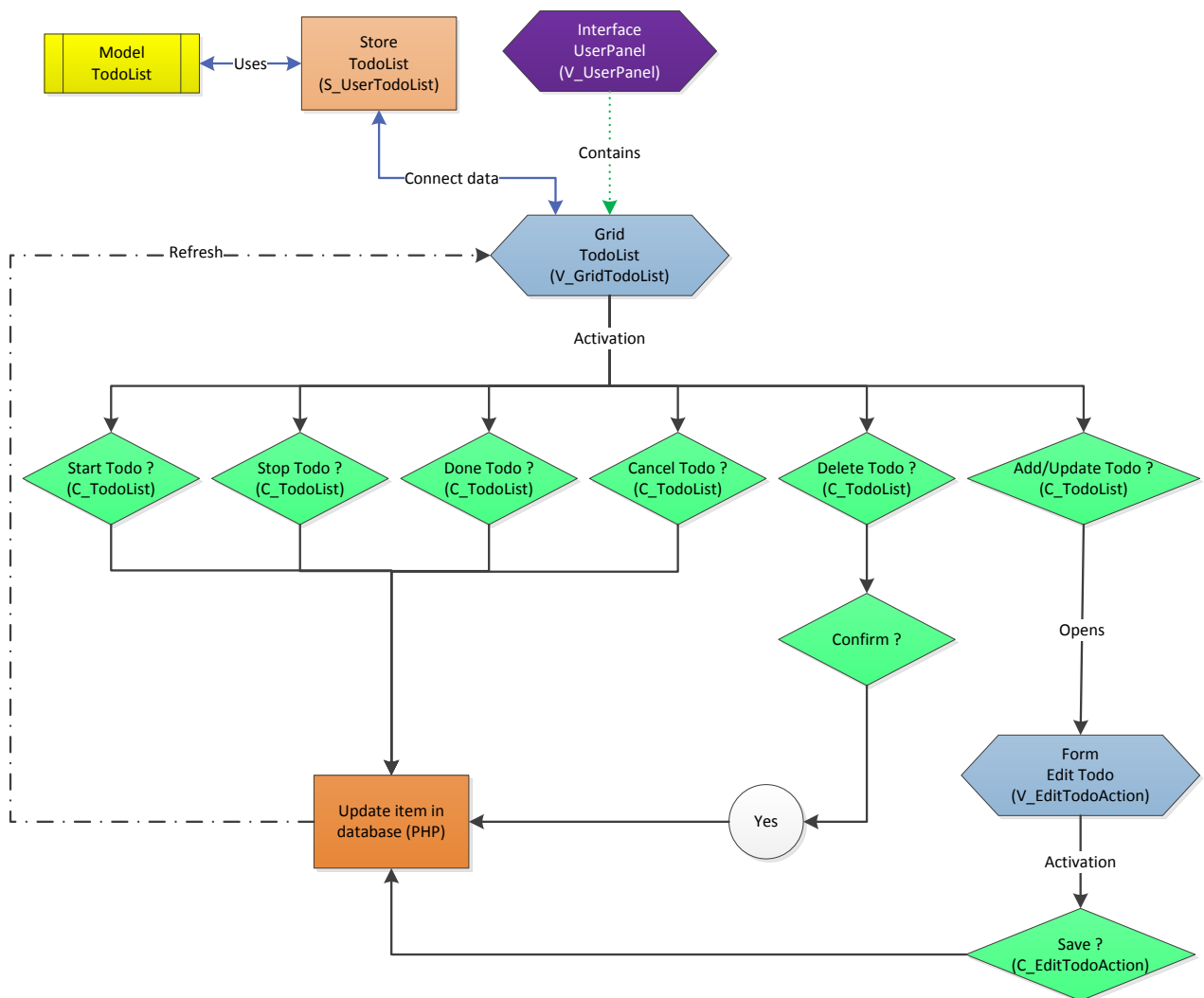


Figure S3. Fonctionnement de la TODO List. Description des différentes actions de la TODO list et leurs conséquences. Entre parenthèse est indiquée la Classe ExtJS concernée. En violet les super-vues contenant d'autres vues, en bleu les vues, en vert les contrôleurs et en orange les scripts PHP. En traits pointillés les actions asynchrones, réalisées une fois la réponse PHP reçue.

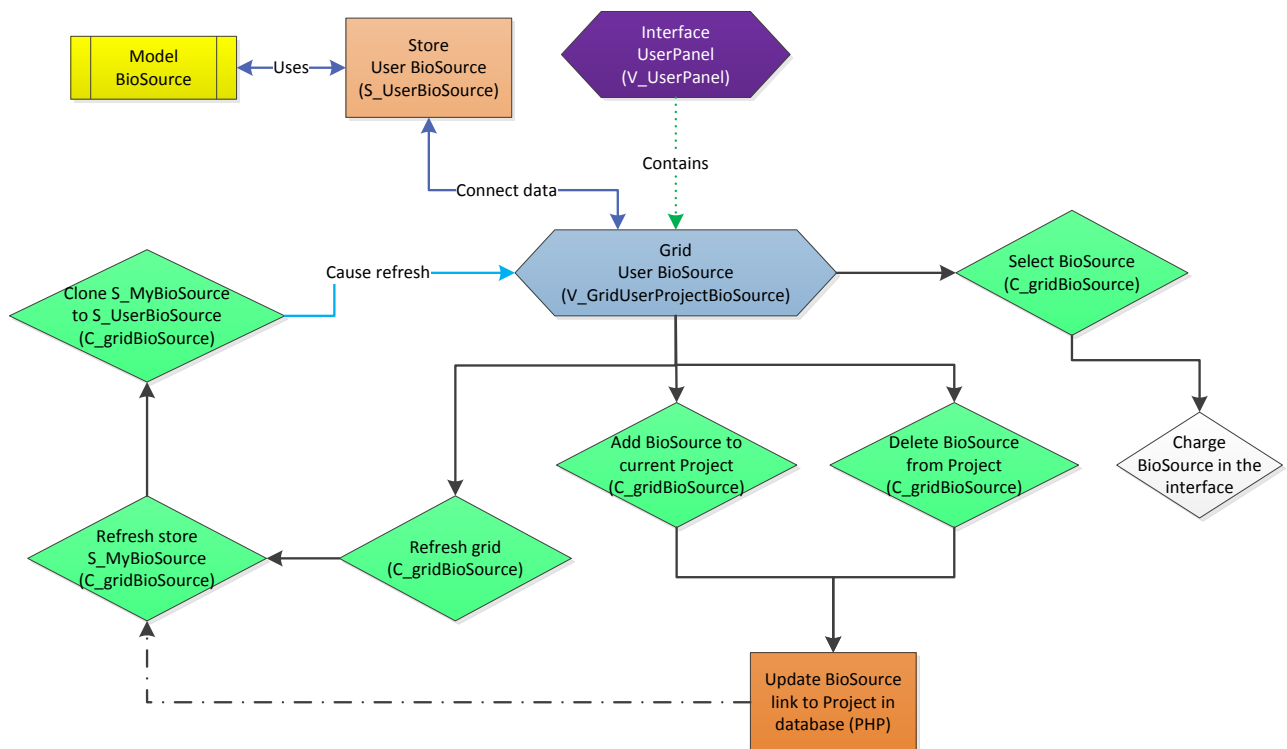


Figure S4. Fonctionnement de la grille *BioSource* Utilisateur. Description des différentes actions sur la grille et leur conséquences. Entre parenthèse est indiquée la Classe ExtJS concernée. En violet les super-vues contenant d'autres vues, en bleu les vues, en vert les contrôleurs et en violet les scripts PHP. En traits pointillés les actions asynchrones, réalisées une fois la réponse PHP reçue. Les losanges correspondent aux actions réalisées côté client.

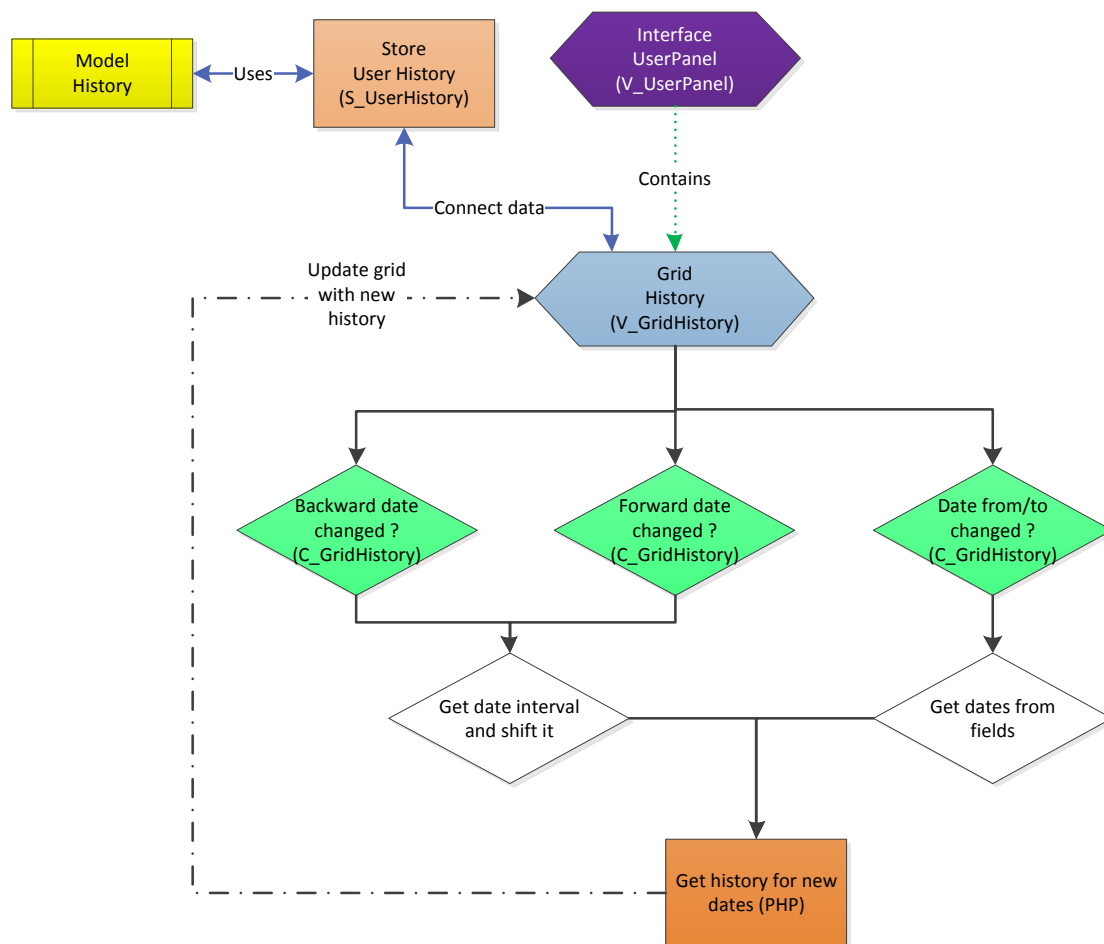


Figure S5. Fonctionnement de la grille Historique. Description des différentes actions sur la grille et leurs conséquences. Entre parenthèse est indiquée la Classe ExtJS concernée. En violet les super-vues contenant d'autres vues, en bleu les vues, en vert les contrôleurs et en orange les scripts PHP. En traits pointillés les actions asynchrones, réalisées une fois la réponse PHP reçue. Les losanges correspondent aux actions réalisées côté client.

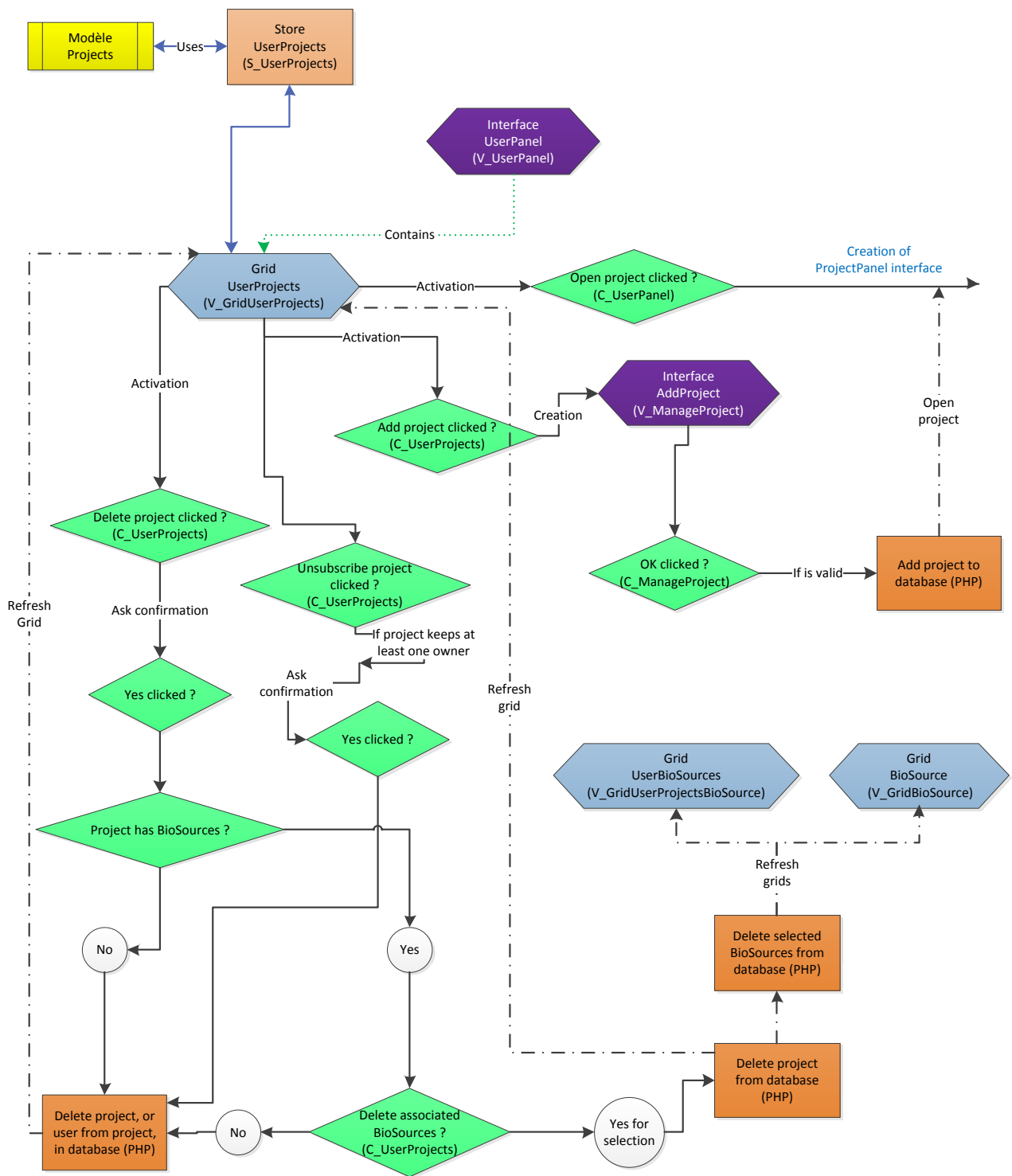


Figure S6 Fonctionnement de la grille Liste des Projets. Description des différentes actions sur la grille et leurs conséquences. Entre parenthèse est indiquée la Classe ExtJS concernée. En violet les super-vues contenant d'autres vues, en bleu les vues, en vert les contrôleurs et en orange les scripts PHP. En traits pointillés les actions asynchrones, réalisées une fois la réponse PHP reçue. Les losanges correspondent aux actions réalisées côté client.

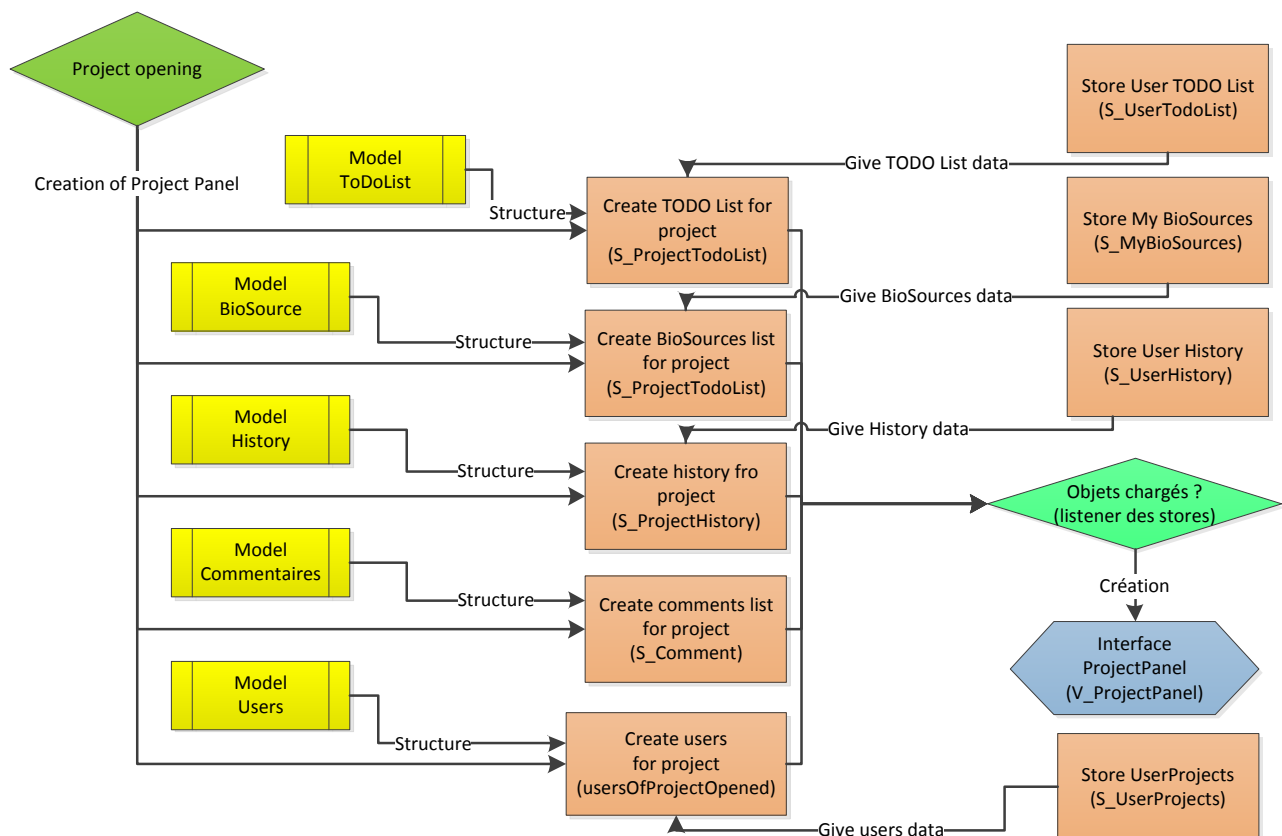


Figure S7. Ouverture d'un *Projet* : initialisation du panneau *Projet*. Entre parenthèse est indiquée la Classe ExtJS concernée. En violet les super-vues contenant d'autres vues, en bleu les vues, en vert les contrôleurs et en marron les stores. Le Store S_Comment est rempli depuis la base de données du serveur via un script PHP. Les autres Stores du panneau sont remplis depuis d'autres Stores déjà chargés, mentionnés sur la figure.

UserPanelProjectPanelNetwork DataNetwork VizNetwork Curation

Project name here

Created 17/01/2015

Edit project

Here description *Altera sententia est, quae definit amicitiam paribus officiis ac voluntatibus. Hoc quidem est nimis exigue et exiliter ad calculos vocare amicitiam, ut par sit ratio acceptorum et datorum. Divitior mihi et affluentior videtur esse vera amicitia nec observare restricte, ne plus reddat quam acceperit; neque enim verendum est, ne quid excidat, aut ne quid in terram defluat, aut ne plus aequo quid in... [see more](#)*

TODO list:

Description	User	Limit date	Status
Private			
Public			

✓

+

-

Comments:

N°	User	Title	type	Attachments
1			Biblio	None
2			Data	see
3			Comment	None
4				
5				
6				
7				
8				
9				

Open

Add new

Delete

History:

Personal

All

Date	User	Action

Personal

All

Figure S8. Schéma du Panneau *Projet* prévue avant l'implémentation concrète dans *MetExplore*.

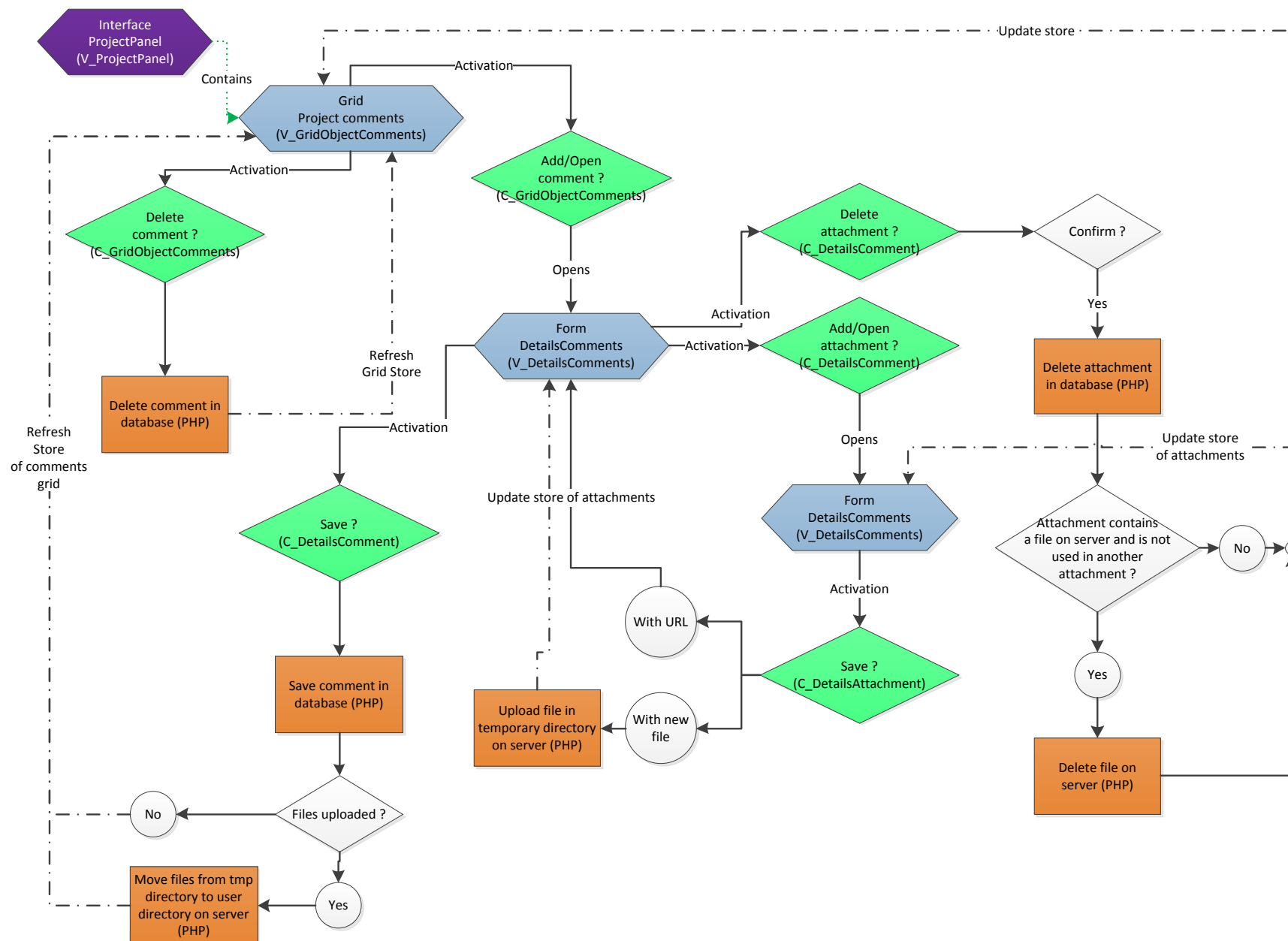


Figure S9. Fonctionnement de la grille Commentaires. Description des différentes actions sur la grille et leurs conséquences. Entre parenthèse est indiquée la Classe ExtJS concernée. En violet les super-vues contenant d'autres vues, en bleu les vues, en vert les contrôleurs et en violet les scripts PHP. En traits pointillés les actions asynchrones, réalisées une fois la réponse PHP reçue.

MetExplore vTest

About Mapping Logout Excel Flux Import Export

User profile Project Details **Network Data** Network Viz Network Curation

BioSources Compartments (1/1) Pathways (377/377) **Reactions (1740/1740)** Metabolites (1810/1810) Enzymatic Complexes (1253/1253)

+ Add Edit Delete Commit Changes Multiple affectation Curation Votes

	Name	Identifier	E.C.	Reversible	Flux Lower B	Flux Upper B
1	α-Arabinofuranosidase	3.2.1.55-RXN	3.2.1.55	☐	0	99999
2	α-amylase	ALPHA-AMYL-RXN	3.2.1.1	☒	-99999	99999
3	α-galactosidase	ALPHAGALACTOSID-RXN	3.2.1.22	☐	0	99999
4	α-glucosidase	MALTODEXGLUCOSID-RXN	3.2.1.20	☒	-99999	99999
5	β-Acetylhexosaminidase	3.2.1.52-RXN	3.2.1.52	☒	-99999	99999
6	β-fructofuranosidase	3.2.1.26-RXN	3.2.1.26	☐	0	99999
7	β-glucosidase	3.2.1.21-RXN	3.2.1.21	☒	-99999	99999
8	β-glucuronidase	BETA-GLUCURONID-RXN	3.2.1.31	☐	0	99999
9	β-ketoacyl-acyl-carrier-protein synthase I	2.3.1.41-RXN	2.3.1.41	☒	-99999	99999
10	β-ketoacyl-acyl-carrier-protein synthase I	RXN-9535	2.3.1.41	☐	0	99999
11	β-ketoacyl-acyl-carrier-protein synthase II	2.3.1.179-RXN	2.3.1.179	☐	0	99999
12	β-ketoacyl-acyl-carrier-protein synthase III	2.3.1.180-RXN	2.3.1.180	☐	0	99999
13	β-lactamase	BETA-LACTAMASE-RXN	3.5.2.6	☒	-99999	99999
14	β-lysine 5,6-aminomutase	BETA-LYSINE-56-AMINOMUTASE-RXN	5.4.3.3	☐	0	99999
15	β-mannosidase	3.2.1.25-RXN	3.2.1.25	☒	-99999	99999
16	β-phosphoglucomutase	BETA-PHOSPHOGLUCOMUTASE-RXN	5.4.2.6	☐	0	99999
17	β-ureidopropionase	BETA-UREIDOPROPIONASE-RXN	3.5.1.6	☐	0	99999
18	γ-glutamylcyclotransferase	GAMMA-GLUTAMLYCYCLOTRANSFE...	2.3.2.4	☐	0	99999
19	γ-glutamyltransferase	GAMMA-GLUTAMYLTRANSFERASE-R...	2.3.2.2	☒	-99999	99999
20	(1→4)-α-D-glucan 1-α-D-glucosylmutase	5.4.99.15-RXN	5.4.99.15	☐	0	99999
21	(2,3-dihydroxybenzoyl)adenylate synthase	DHBAMPLIG-RXN	2.7.7.58	☐	0	99999
22	(2E,6E)-farnesyl diphosphate synthase	FPPSYN-RXN	2.5.1.10	☐	0	99999
23	(2R)-sulfolactate sulfo-lyase	RXN-11691	4.4.1.24	☐	0	99999
24	(2E)-3-[(3aS,4S,5R,7aS)-5-hydroxy-7a-methyl-1-oxo-oct...	RXN-12749	4.2.1.-	☒	0	99999

Selected BioSource

Public: -- Select public BioSource --

Private: -- Select private BioSource --

Project: Agrobacterium tumefaciens (Strain: 5A v2) (Source: Agromics, Version: 0)

Agrobacterium tumefaciens (Strain: 5A v2) (Source: Agromics, Version: ...

Save Reset + Add Publication Share Delete Duplicate

Comp	Path	Rxn	Met	Enz	Prot	Genes
1	377	1740	1810	1253	1253	1245

BioSource Data

Name: AGRT5A

Organism: Agrobacterium tumefaciens

Strain/ Cell Line : 5A v2

Tissue:

Cell Type:

Source Database: Agromics

URL: http://microcyc.genoscope.cns.fr/

Id in Database: AGRT5A1328

Version: 08-07-2014

Cart +

Jobs +

Figure S10. Interface du Serveur Web MetExplore. Le panneau principal de MetExplore comporte 5 onglets. Les deux premiers permettent d'accéder respectivement au panneau Utilisateur et au panneau Projet. Le troisième, « Network Data », permet de visualiser les données du réseau dans des grilles. Ainsi, à l'intérieur de cet onglet sont présents dans le premier onglet la liste des BioSources accessibles à l'utilisateur connecté. Les autres onglets affichent les composants du réseau contenu dans la BioSource sélectionnée : les compartiments cellulaires, les voies métaboliques (« Pathways »), les réactions, les métabolites, les complexes enzymatiques, les protéines et les gènes. Pour chacun de ces types d'objets du réseau, l'onglet contient une liste des objets contenus dans le réseau dans une grille, avec plusieurs de ces caractéristiques comme le nom et l'identifiant de chaque objet. Le quatrième onglet (« Network Viz ») propose une visualisation de tout ou partie du réseau métabolique sous forme d'un graphe. Le dernier onglet propose des formulaires de modification, ou d'ajouts de nouveaux objets au réseau. La barre d'outil du haut permet de réaliser divers mappings ou analyses de flux, l'import/export de réseaux métaboliques entre autre au format SMTL et Excel (pour l'export uniquement pour ce dernier), et de se (dé)connecter. Enfin, Le panneau de droite permet de sélectionner une BioSource, de voir ou modifier les informations de la BioSource sélectionnée, d'ajouter des réactions au panier en vue de leur visualisation dans le panneau de visualisation ou enfin de visualiser les longues tâches en cours.

DESCRIPTION DETAILLE DES PRINCIPAUX OUTILS DE RECONSTRUCTION DE LA LIT- TERATURE

Plusieurs outils de reconstructions existent. KAAS (Moriya *et al*, 2007) est un outil en ligne permettant de générer une reconstruction métabolique automatique à partir des informations génomiques. Il recherche pour cela les gènes orthologues avec d'autres espèces proches dont la reconstruction est déjà effectuée pour déduire la fonction de chaque gène et le placer dans le nouveau réseau. Il considère cependant comme orthologue 2 gènes de 2 espèces distinctes qui sont des *Bidirectional Best Hit* (BBH), c'est-à-dire que pour le gène A de l'espèce 1, le gène qui lui est le plus similaire chez l'espèce 2 est le gène B, et le gène le plus similaire au gène B chez l'espèce 1 est le gène A. Ceci n'est pas la définition rigoureuse d'un gène orthologue.

Pathway tools (Karp *et al*, 2002) permet la visualisation du réseau métabolique, la reconstruction d'un réseau métabolique à partir des annotations génomiques et l'édition de la reconstruction afin d'y apporter des corrections manuelles. Il dispose d'un historique pour chaque reconstruction de toutes les modifications apportées. Ainsi, pour un projet à plusieurs chacun peut savoir qui a modifié quoi et quand. Plusieurs utilisateurs peuvent faire des modifications, le réseau étant stocké sur un serveur. Pour prédire un nouveau réseau, le logiciel se sert des réseaux déjà existants dans la base de données MetaCyc, par comparaison des nouveaux éléments à ceux de ces réseaux. Pour cela, pour chaque voie métabolique P des reconstructions existantes, le logiciel évalue l'évidence qu'elle existe dans le réseau de l'organisme S à reconstruire. Cette évidence est basée sur le nombre d'enzymes de la voie métabolique P qui sont retrouvées chez l'organisme S, selon l'annotation fonctionnelle de S fournie. Il est possible de modifier au préalable les annotations du génome dans le logiciel si elles contiennent des erreurs. Une fois le réseau reconstruit, il est également proposé un éditeur permettant de le corriger.

rBioNet (Thorleifsson & Thiele, 2011) est un outil ayant pour objectif la reconstruction métabolique de qualité (contrôle et assurance qualité [QC/QA]). C'est un outil que l'on utilise dans Matlab, aux côtés de plusieurs mesures QC/QA déjà présentes dans la COBRA Toolbox (Becker *et al*, 2007). rBioNet est constitué de 3 parties : un créateur de métabolite, un créateur de réaction et un créateur de reconstruction. Pour assurer une bonne qualité du réseau, les métabolites impliqués dans une réaction doivent nécessairement exister dans la base de données des métabolites. Le créateur de reconstruction métabolique permet de créer une reconstruction à partir de rien, ou d'en importer une (depuis le format Matlab, une feuille de calcul, ou un fichier SBML), pour la compléter ou la corriger. Encore une fois, seules les réactions présentes dans la base de données des réactions peuvent être ajoutées à la reconstruction. Aussi, chaque réaction reconstruite peut être associée à des

gènes via l'association gène-protéine-réaction (GPR). Le modèle reconstruit peut ensuite être exporté au format Matlab ou SBML. Cet outil ne peut s'utiliser que localement.

GEMSiRV (Liao *et al*, 2012) est un outil qui a pour but de proposer à la fois de la reconstruction de réseau métabolique et divers simulations sur le réseau reconstruit. Les outils de reconstruction proposés permettent de reconstruire automatiquement un réseau métabolique brouillon grâce à un réseau déjà reconstruit pour une espèce voisine, où les réactions sont connectées en comparant les objets des 2 réseaux, déjà reconstruit et à reconstruire. Il est également possible d'importer un réseau existant depuis un fichier SBML ou Excel. Ensuite, dans tous les cas, il est possible via l'interface de corriger le réseau manuellement. Cet outil ne peut s'utiliser que localement.

PathwayBooster (Liberal *et al*, 2015) est un outil ayant pour objectif la comparaison et la correction de réseaux métaboliques. Il permet de comparer différents réseaux métaboliques. Il est connecté à la base de données KEGG. Il peut représenter les réactions d'un réseau KEGG avec un code couleur permettant de voir si elle existe chez l'organisme étudié. L'outil permet de visualiser le réseau et d'identifier les erreurs, qui devront être corrigées ailleurs, l'outil ne proposant pas d'édition du réseau.

RAVEN Toolbox (Agren *et al*, 2013) est un outil de reconstruction, d'analyse et de visualisation de réseaux métaboliques. Il permet de reconstruire un réseau à partir d'un réseau déjà existant, grâce aux relations d'orthologie. Et il permet de combler les vides grâce aux *KEGG Orthology* (KO). Il s'exécute à l'intérieur de Matlab.

Enfin, parmi les principaux outils, reste Model SEED (Henry *et al*, 2010) qui permet de générer un modèle métabolique automatiquement en 48 heures à partir d'une séquence génomique. Une fois la séquence génomique annotée automatiquement, il est recommandé d'utiliser d'autres outils afin de faire les corrections manuelles nécessaires, avant de se lancer dans la reconstruction automatique du réseau. Cette reconstruction se fait automatiquement en suivant un pipeline précis, et des analyses sont notamment faites pour tester et classifier les réactions (Flux Variability Analysis [FVA], Flux Balance Analysis [FBA], ...). On obtient un modèle assez précis mais des corrections manuelles restent nécessaires, mais ne sont pas proposées directement via cet outil.