

Travaux pratiques Formation

OCTOBRE 2016

Les pipelines sont accessibles depuis l'adresse :

<https://bbric-pipelines.toulouse.inra.fr/galaxy/>

Transcriptomique

TP Assemblage Transcriptome

bbric.toulouse.inra.fr, <https://bbric-pipelines.toulouse.inra.fr/galaxy>

1- Récupérer les données dans Galaxy

On souhaite reconstruire le transcriptome de la bactérie *S bbric*.

Dans Galaxy, utilisez l'outil BBRIC Archive pour récupérer les données RNASeq de cette bactérie (3bibliothèques pour 5 fichiers). Profitez en pour regarder leurs caractéristiques en lisant les métadonnées.

2- Lancement de l'assemblage du Transcriptome

Accéder à l'outil d'assemblage de transcriptome, soit à partir de bbric.toulouse.inra.fr, soit en navigant dans Galaxy (tools BBRIC Protocols)

Régler les paramètres en fonction des bibliothèques que vous souhaitez assembler (gamme de K-mer)

→ Faire tourner le pipeline avec 2 jeux de paramètres différents

3- Analyse des fichiers de résultats :

Fichier de stats : Quelles différences constatez-vous et comment l'expliquez vous (stringence, taille de K-mer) ?

Fichier fasta : Que feriez vous pour estimer la qualité de votre assemblage ?

TP Mesure de l'expression

bbric.toulouse.inra.fr, <https://bbric-pipelines.toulouse.inra.fr/galaxy>

1- Récupérer les données dans Galaxy

On souhaite étudier l'expression des gènes de la bactérie *S bbric*.

Dans Galaxy, allez dans le menu Shared Data et sélectionnez "Data Libraries". Cliquez sur "BBRIC Protocols". Dans la partie "S bbric", sélectionnez les données souhaitées :

-- La séquence génomique et son annotation :

- * S. bbric genomic sequence
- * S. bbric structural annotation

-- **Librairies RNAseq (au choix) :**

Librairie Sb-2404-33-R1 (single end orienté) (50nt)

*S.bbric-RbmLong-GGK33.fastq.gz

ou

Librairie Sb-2404-21-R1 (paired end orienté, le read 1 qui donne l'orientation)

*S.bbric-RbmLong-GGK21.ope.1.fastq.gz

*S.bbric-RbmLong-GGK21.ope.2.fastq.gz

ou

Librairie Sb-2404-36-R1 (paired end orienté, le read 1 donne l'orientation, enrichie en petits ARN)

* S.bbric-RbmSmall-GGK36.ope.1.fastq.gz

* S.bbric-RbmSmall-GGK36.ope.2.fastq.gz

Sélectionnez "Import to current history" et cliquez sur GO : les fichiers d'entrée pour tester ce pipeline sont maintenant dans votre historique.

2- Lancement de la mesure de l'expression

Accéder à l'outil de mesure de l'expression, soit à partir de bbric.toulouse.inra.fr, soit en navigant dans Galaxy (tools BBRIC Protocols)

Régler les paramètres en fonction des librairies que vous souhaitez analyser :

Maximal distance between paired reads (si pair) = 70nt pour les Small ; 130nt pour les Long

→ Faire tourner le pipeline avec 2 valeurs différentes pour le nombre de mismatches autorisé (0 et 2)

3- Analyse des fichiers de résultats :

Fichier de stats : Quelle différence constatez-vous entre 0 et 2 mismatches ?

Fichier de count : Quelle différence constatez-vous entre 0 et 2 mismatches ? Est-ce ce à quoi vous attendiez ?

Analyse du polymorphisme

Détection des SNPs (samtools/VarScan) pour analyse de leurs effets

1/Récupération du jeu de test dans share data

Shared Data ☺ Data Librairies ☺ BBRIC protocols > cocher 'X.bbric' (sélectionne toutes les données)

☺ 'Import to current history' Go

Xbbric_A = séquences de l'échantillon A (2 fastq, 100nt)

Xbbric_B = séquences de l'échantillon B (2 fastq, 100nt)

Xbbric_C = séquences de l'échantillon C (2 fastq, 76nt)

Xbbric genomic sequence = séquence du génome (fasta)

Xbbric structural annotation = annotation du génome (gff)

2/Charger le workflow d'analyse RADseq avec STACKS

BBRIC protocols ☺ POLYMORPHIMS > SNP detection and effect prediction

Paramétrer l'analyse

- Cocher 'Paired-end library'
- 'Maximum distance between paired reads (nt)' : 300
- Cliquer 2 fois sur 'Add new Paired-end library'
- Sélectionner les reads des 3 individus et mettre A, B, et C dans les 'Sample name'
- Sélectionner 'Xbbric genomic sequence' pour 'Reference genome file (fasta)'
- Sélectionner 'Xbbric structural annotation' pour 'Genome annotation file (GFF3)'
- 'Minimum hit length' :50 (car échantillon C a une longueur de reads de 76nt)
- 'Maximum number of mismatches': 4
- 'Minimum position coverage': 20
- 'Minimum variant coverage': 10
- 'Minimum position frequency': 0.2

- 'Minimum homozygous frequency': 0,75
- 'Effect results filter': Select All
- Cocher 'Upper cas nucleotides'

3/Analyse des résultats

Nous obtenons trois fichiers de résultats :

SNP effect by gene : Nombre de chaque catégorie d'effets trouvés pour chaque gène de l'annotation

SNP effect report : Rapport HTML produit par SnpEff avec des statistiques sur le type et le nombre d'effets prédits.

SNP matrix file : Fichier VCF zippé

PacBio data analyses

**TP « ANALYSE DU METHYLOME BACTERIEN AVEC DES DONNEES PacBio
ET L'ENVIRONNEMENT D'ANALYSE « SMRT portal » »**

Partie 1 : Analyse des données SMRT pour la détection des méthylations

1. Ouvrir dans votre navigateur <https://bbric-pipelines.toulouse.inra.fr/smrtanalysis>
2. Identifiez-vous avec vos identifiants « authentification fédération » (votre ldap national)
3. Vous allez recevoir un email ayant pour objet : « [BBRIC] SMRT Analysis account »
Celui-ci contiendra votre login/password
Connectez-vous
4. Cliquez sur « **Create New** »
 - Sélectionnez l'échantillon **Xbbric_949** disponible dans la partie à gauche « SMRT cells Available » et faites le passer à l'aide de la flèche dans la « SMRT Cells in Job » à droite
 - Sélectionnez dans la liste Protocol : **RS_Modification_and_Motif_Analysis.1**
 - Sélectionnez Reference : **Xbbric**
 - Donnez un nom à votre Job dans le champ « Job Name »
 - Sélectionnez le « Groups » (public : ouvert à tous, votre_login : privé)
 - ⇒ Cliquez sur Save
 - ⇒ Cliquez sur Start
5. Votre job terminé, vous allez recevoir un email de SMRT Portal
 - => Récupérez le JobId de votre analyse : « _____ »
 - => Si votre run n'est pas terminé, vous pouvez utiliser ceux-ci : « 16547 » ou « 16551 »
 - a. La couverture de la séquence vous semble suffisante ?
 - b. Combien de type de méthylation sont identifiées ?
 - c. Allez sur <http://rebase.neb.com/rebase/rebase.html> et cherchez un motif identifié en sélectionnant 'recognition sequence'.

Partie 2 : Détection des motifs méthylés

1. Accédez au pipeline « **Methylation motifs detection** » via le moteur de recherche accessible sur le portail du CATI-BBRIC.
2. Paramétrez l'outil en renseignant le numéro « JobId » obtenu en sortie de SMRT portal, ainsi que
 - le champ "Minimum QV score to use in motif finding "
 - et le champ "Only use motifs above this methylated fraction"
3. Lancez votre analyse.

- a. Que vous indique la ligne Not clustered dans le fichier « Methylated motifs » ? Est-ce correct ?
- b. Quels types de modifications avez-vous ?
- c. Quels motifs ont disparu ?
- d. Le pourcentage de motifs détectés a-t-il changé ?

Partie 3 : Visualisation via myGenomeBrowser

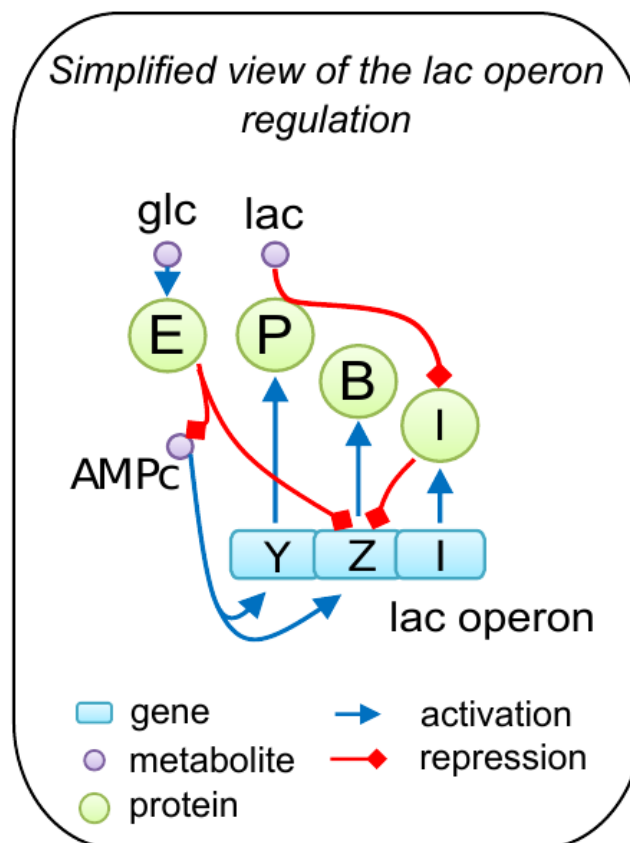
1. Accédez au pipeline « Methylation visualization » via le moteur de recherche accessible sur le portail du CATI-BBRIC.
2. Récupéré l'annotation structurale de XbbriC dans le menu « Shared Data » puis « XbbriC ». Cocher le fichier « XbbriC structural annotation » et « Go »
3. Paramétrez l'outil en renseignant le numéro « JobId » obtenu en sortie de SMRT portal, ainsi que les autres champs si vous disposez des fichiers demandés.
Vous pouvez aussi mettre plusieurs « JobId » pour comparer (16547, 16551).
4. Vous pouvez observer le résultat dans « Methylation data ».
 - a. Les méthylations sont majoritairement dans des gènes ou dans des régions inter-géniques ? Lesquels sont potentiellement les plus intéressants ?
 - b. Il y a 2 paires de motifs, trouvez des hémi-méthylations.
5. Comparer les 2 JobId précédents. Observez-vous des différences entre les 2 ?
6. Téléchargez le zip résultat.
7. Décompressez-le.
8. Ouvrez dans votre navigateur préféré « myGenomeBrowser »
URL : <https://bbriC-pipelines.toulouse.inra.fr/myGenomeBrowser>
- NB : Ici utilisez l'URL suivante, le génome de référence est déjà chargé pour vous :
<https://bbriC-pipelines.toulouse.inra.fr/myGenomeBrowser?browse=1&portalname=XbbriC&owner=Sebastien.Carrere@toulouse.inra.fr&key=8l2uoPKX>
9. cliquez sur File>Open>Local files> et sélectionner les fichiers
10. Sélectionner tous les fichiers du zip, puis faites open en bas.

Naviguez dans vos résultats

Visualisation de réseaux avec Cytoscape

Objectifs :

- Créer un réseau et le charger dans Cytoscape
- Créer une table d'attributs et la charger dans Cytoscape
- Changer les caractéristiques des nœuds et des arêtes en fonction de leurs attributs
- Réorganiser le dessin pour qu'il soit le plus lisible possible pour obtenir une image ressemblant à celle-ci :



1.1 Ecrire un fichier sif décrivant le réseau ci-dessus.

Un fichier sif est une liste d'arêtes (interactions)¹. Afin de pouvoir changer le style visuel des arêtes en fonction de leur nature, il est important de noter dans le fichier la nature de chaque interaction.

Par exemple, ici :

glc activates E

E inhibits Z

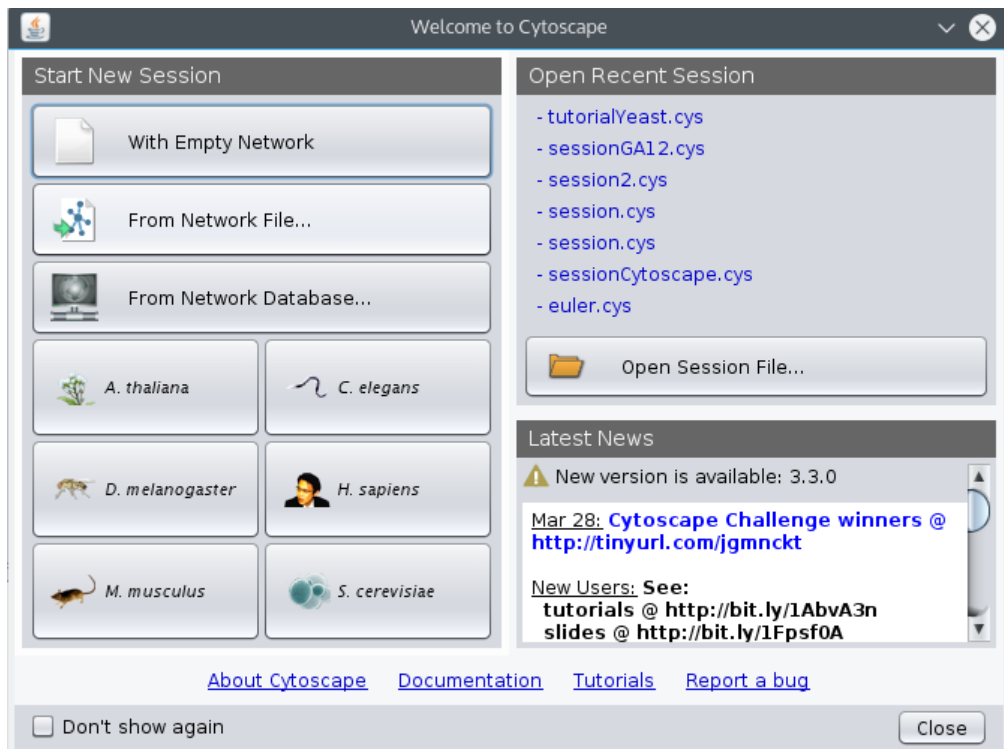
...

1.2 Charger le fichier dans Cytoscape

Ouvrir Cytoscape.

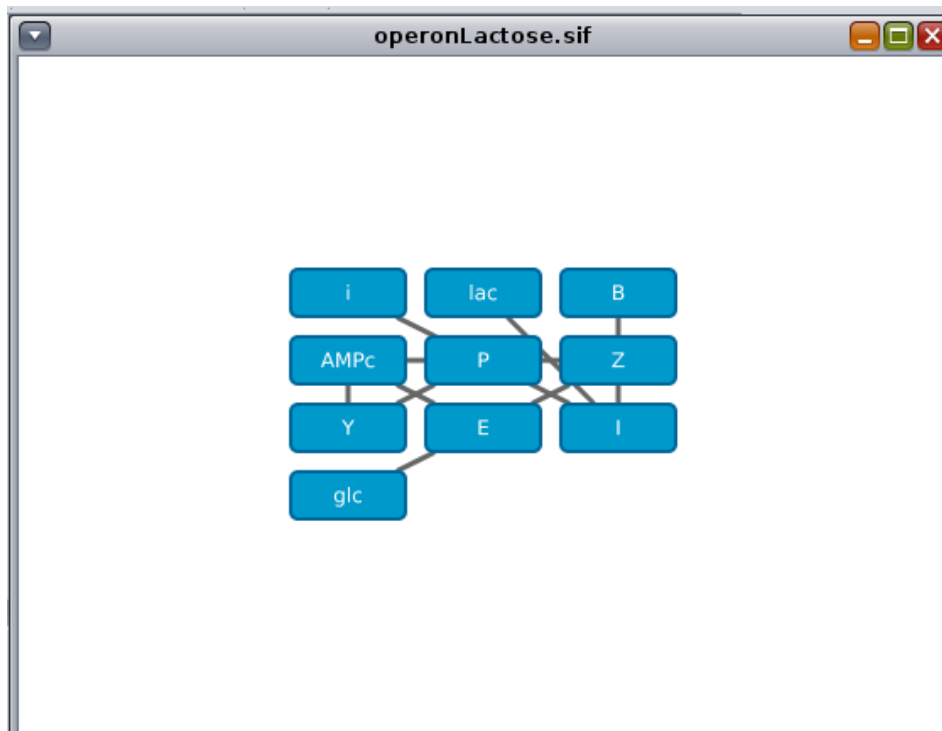
Cliquer sur « From Network File » et choisir le fichier sif.

¹http://wiki.cytoscape.org/Cytoscape_User_Manual/Network_Formats



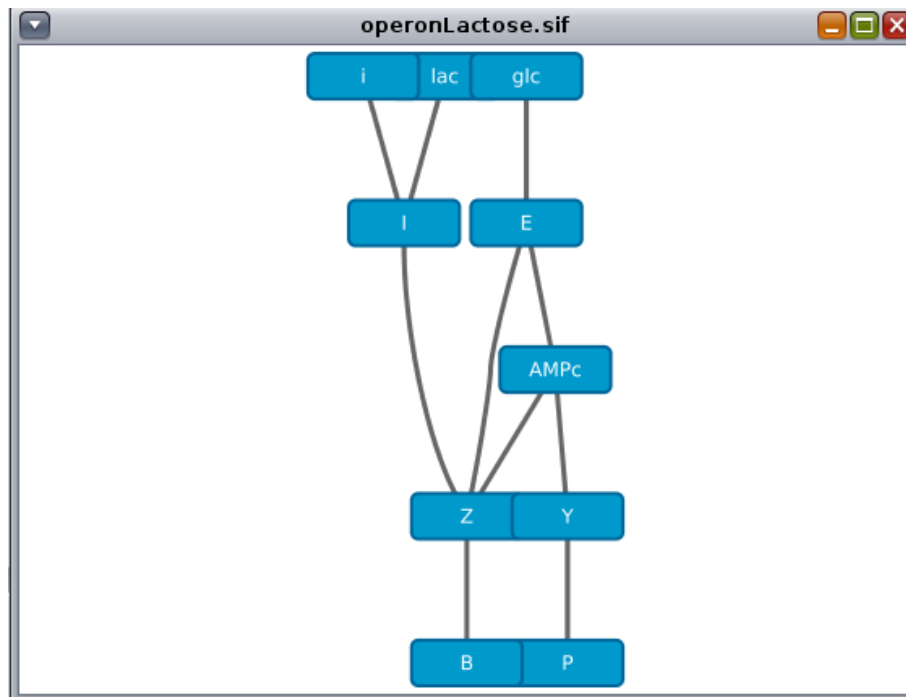
Cliquer ensuite sur Ok pour créer une nouvelle collection de networks.

Vous devez obtenir cette image (ou à peu près) :



Jouez avec les différents layouts (Menu Layout) pour réorganiser automatiquement les nœuds.

Exemple : yfilesLayout → Hierarchic



1.3 Ecrire un fichier d'attributs pour les nœuds du réseau.

Dans ce fichier csv (ou tabulé), nous allons préciser la nature de chaque nœud. Donner un en-tête au fichier pour retrouver facilement vos attributs ensuite dans Cytoscape.

Exemple :

ID,Type

glc,metabolite

Y,gene

...

1.4 Charger le fichier dans Cytoscape

Aller dans Import→Table→File et choisir le fichier créé précédemment

Dans la page d'import, cochez les bons paramètres pour que le délimiteur du fichier soit la virgule et que la première ligne soit considérée comme un en-tête. Cliquer sur Ok.

Import Columns From Table

Target Table Data

Where to Import Table Data: To a Network Collection

Select a Network Collection

Network Collection: operonLactose.sif

Key Column for Network: shared name

Importing Type

Import Data as: Node Table Columns

Advanced

Show Mapping Options
 Show Text File Import Options
 Case Sensitive

Text File Import Options

Delimiter
 Tab
 Comma
 Semicolon
 Space
 Other

Preview Options
 Show all entries in the file

Column Names
 Transfer first line as column names
 Start Import Row: 1
 Comment Line:

Preview

Text File

newTable

ID	Type
glc	metabolite
Y	gene
Z	gene
i	gene
lac	metabolite

Left Click: Enable/Disable Column, Right

OK Cancel

Dans la table en-dessous du réseau, apparaissent maintenant les attributs chargés.

operonLactose.sif

Table Panel

share...	name	Type
	glc	metab...
	E	protein
	lac	metab...
	I	protein
	AMPc	metab...
	Z	gene
	Y	gene
	P	protein
	B	protein

1.5 Sélection de nœuds dans un réseau

Utiliser le champ de recherche pour sélectionner les nœuds correspondant à des gènes.

operonLactose.sif

gene

Table Panel

share...	name	Type
	Z	gene
	Y	gene
	i	gene

Sélectionner les nœuds activés ou réprimés par les gènes sélectionnés.

The screenshot shows a software window titled 'operonLactose.sif' displaying a network diagram of the lac operon. The nodes are arranged in a hierarchical structure: 'i', 'lac', and 'glc' at the top; 'I' and 'E' below them; 'AMPc' to the right of 'E'; 'Z' and 'Y' below 'I' and 'E'; and 'B' and 'P' at the bottom. A context menu is open over the 'I' node, listing various actions such as 'First Neighbors of Selected Nodes', 'Invert node selection', 'Hide selected nodes', 'Show all nodes', 'Select all nodes', 'Deselect all nodes', 'Nodes connected by selected edges', and 'From ID List file...'. A search bar at the top right of the window contains the text 'gene'. The main application window has a menu bar with 'Select', 'Layout', 'Apps', 'Tools', and 'Help'.

1.6 Changement du style des nœuds en fonction des attributs

Nous allons d'abord changer la couleur des nœuds en fonction de leur type.

Cliquer sur l'onglet Style dans le Control Panel et choisissez la propriété Fill Color. Choisissez la colonne « Type » et « Discrete Mapping ». Ensuite choisissez une couleur pour chaque type de nœud.

The screenshot shows the Cytoscape software interface. On the left is the 'Control Panel' with various settings for the network visualization. On the right is a window titled 'operonLactose.sif' displaying a network diagram of the lac operon.

Control Panel Settings:

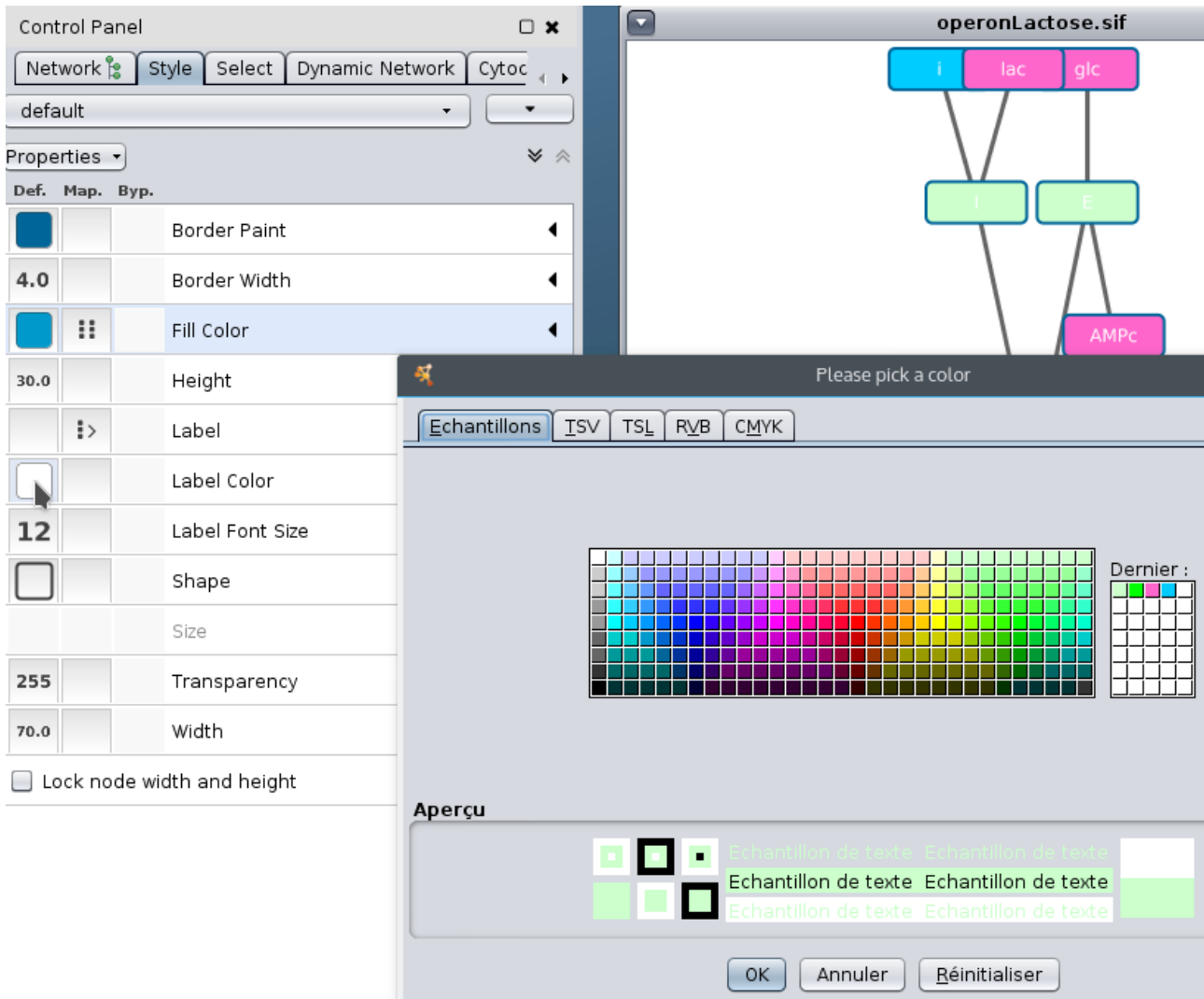
- Network: default
- Properties: Def. Map. Byp.
- Border Paint: [Blue square]
- Border Width: 4.0
- Fill Color: [Blue square]
- Mapping Type: Discrete Mapping
- gene: R:0 G:204 B:255 - #00...
- metabolite: R:255 G:102 B:204 - #...
- protein: R:204 G:255 B:204 - #...
- Height: 30.0
- Label: [Arrow icon]
- Label Color: [White square]
- Label Font Size: 12
- Shape: [Square icon]
- Size: [Info icon]
- Transparency: 255

operonLactose.sif Network Diagram:

```

    graph TD
      i[i] --- E[E]
      lac[lac] --- E
      glc[glc] --- E
      E --- Z[Z]
      E --- Y[Y]
      AMPc[AMPC] --- Z
      AMPc --- Y
      Z --- B[B]
      Y --- P[P]
  
```

Pour certains des nœuds, la couleur rend illisible le label, nous allons donc changer la couleur par défaut des labels. Pour cela, cliquer sur la colonne Def. de la propriété « Label Color » et choisissez Noir.



Control Panel

Network Style Select Dynamic Network Cytoc

default

Properties

Def. Map. Byp.

Border Paint

4.0 Border Width

Fill Color

30.0 Height

Label

Label Color

12 Label Font Size

Shape

Size

255 Transparency

70.0 Width

Lock node width and height

operonLactose.sif

Please pick a color

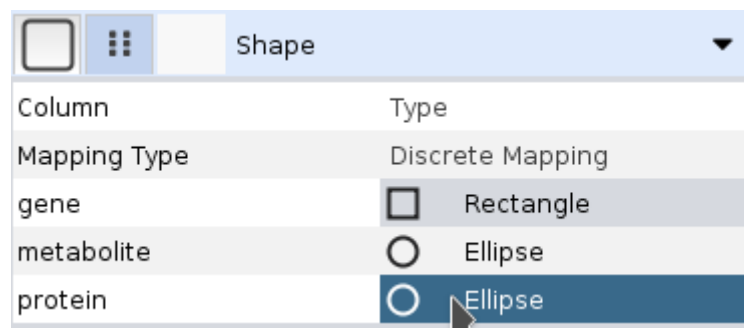
Echantillons ISV TSL RVB CMYK

Dernier :

Aperçu

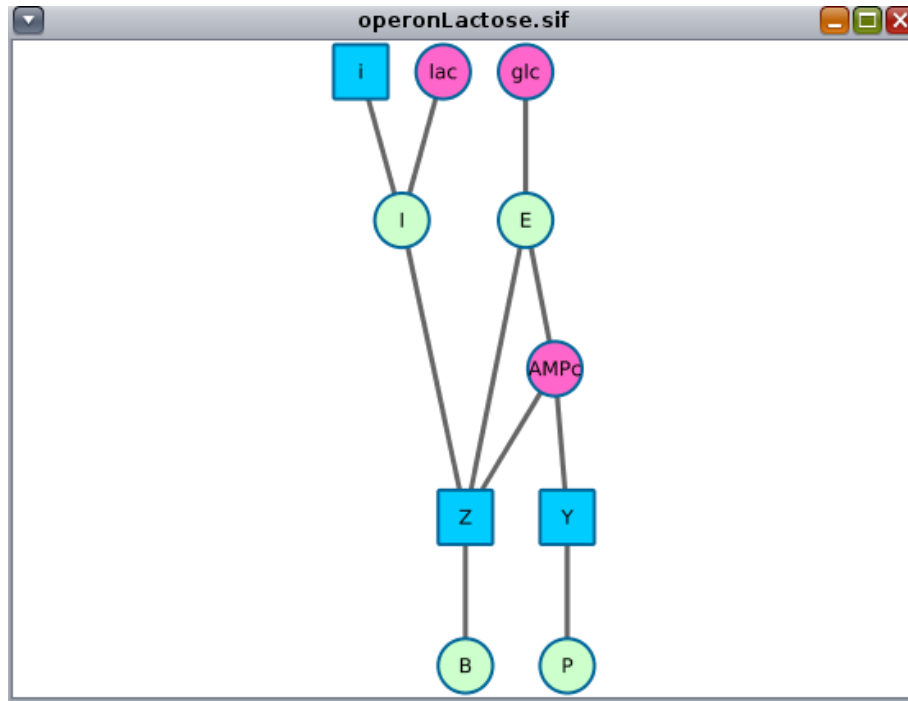
OK Annuler Réinitialiser

Nous allons maintenant changer la forme des nœuds en fonction de leur type.



Column	Type
Mapping Type	Discrete Mapping
gene	<input type="checkbox"/> Rectangle
metabolite	<input type="radio"/> Ellipse
protein	<input checked="" type="radio"/> Ellipse

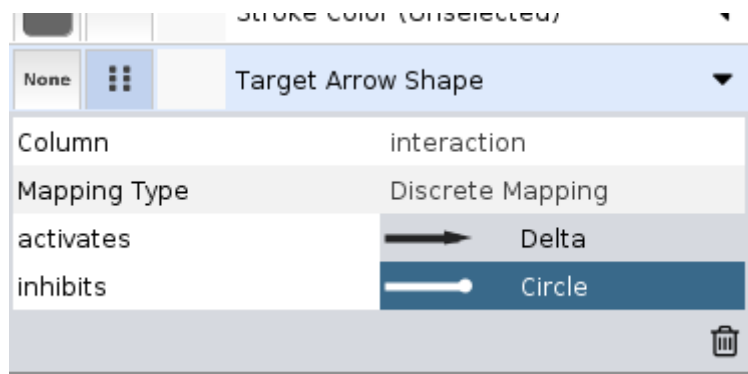
Pour obtenir des cercles, sélectionner « Lock node width and height ».



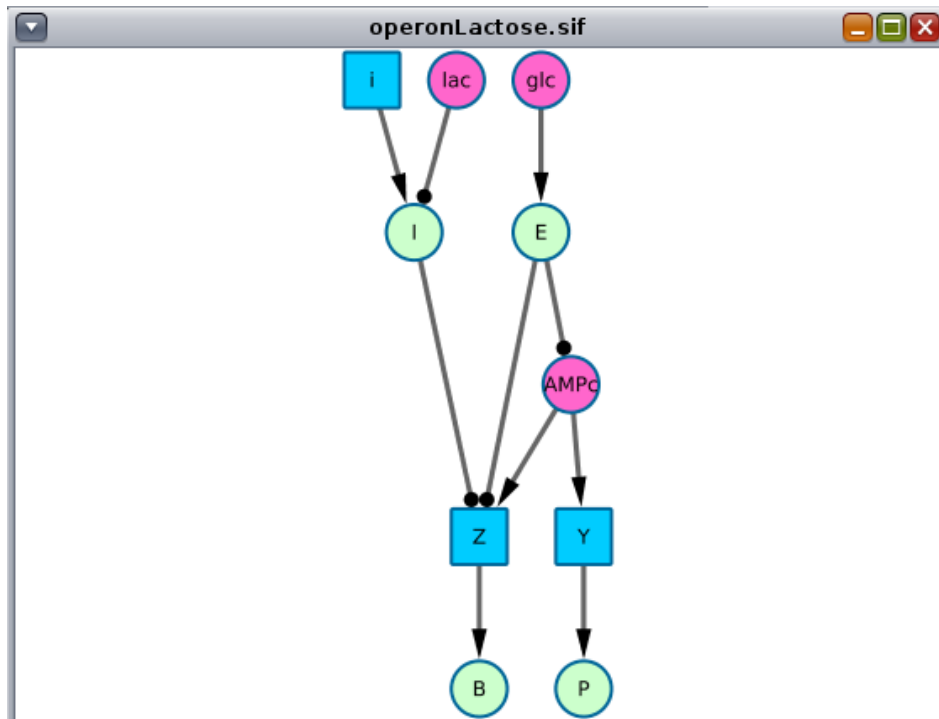
1.6 Changement du style des aretes en fonction de leur nature

Cliquer dans l'onglet Edge dans Style.

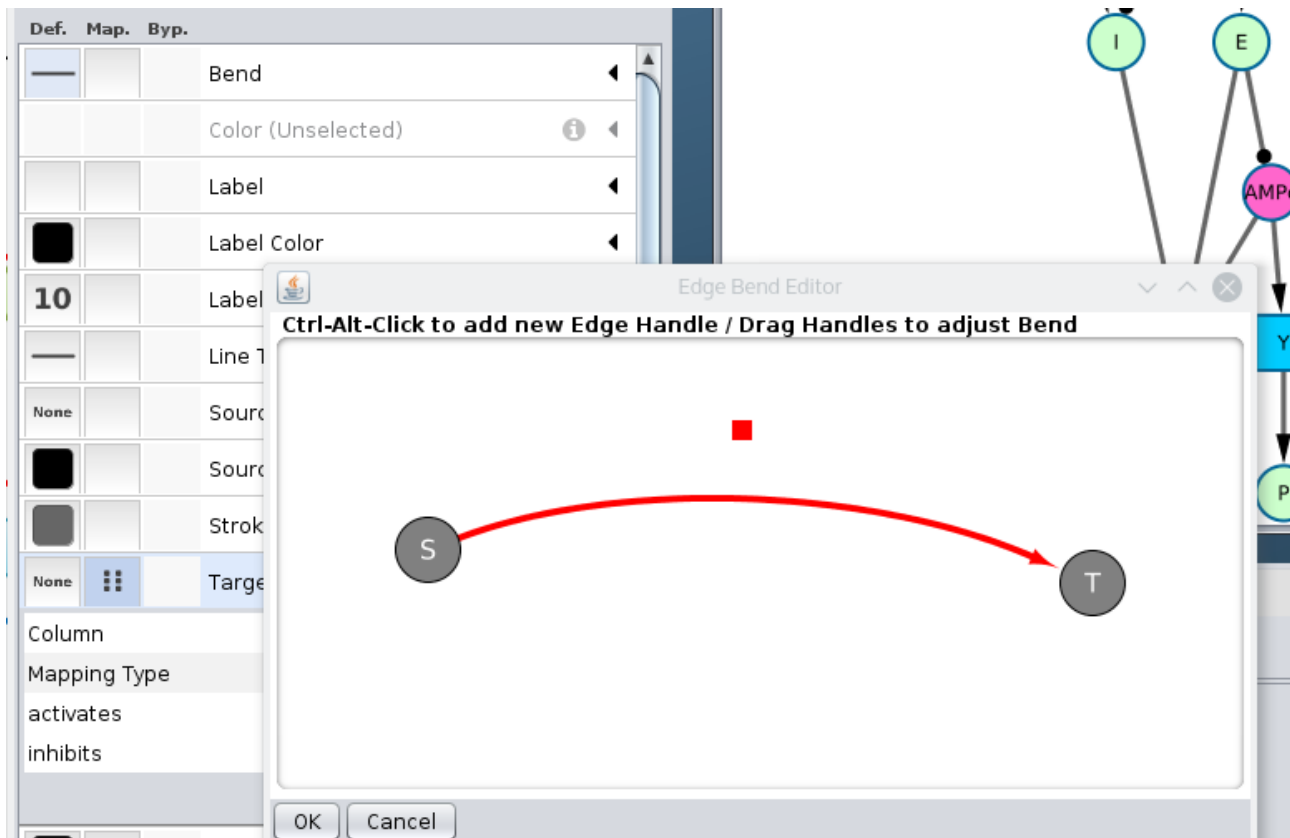
Choisir Target Arrow Shape, la colonne interaction et « Discrete Mapping » et choisir la forme de la flèche pour chaque valeur.



On obtient ce réseau :



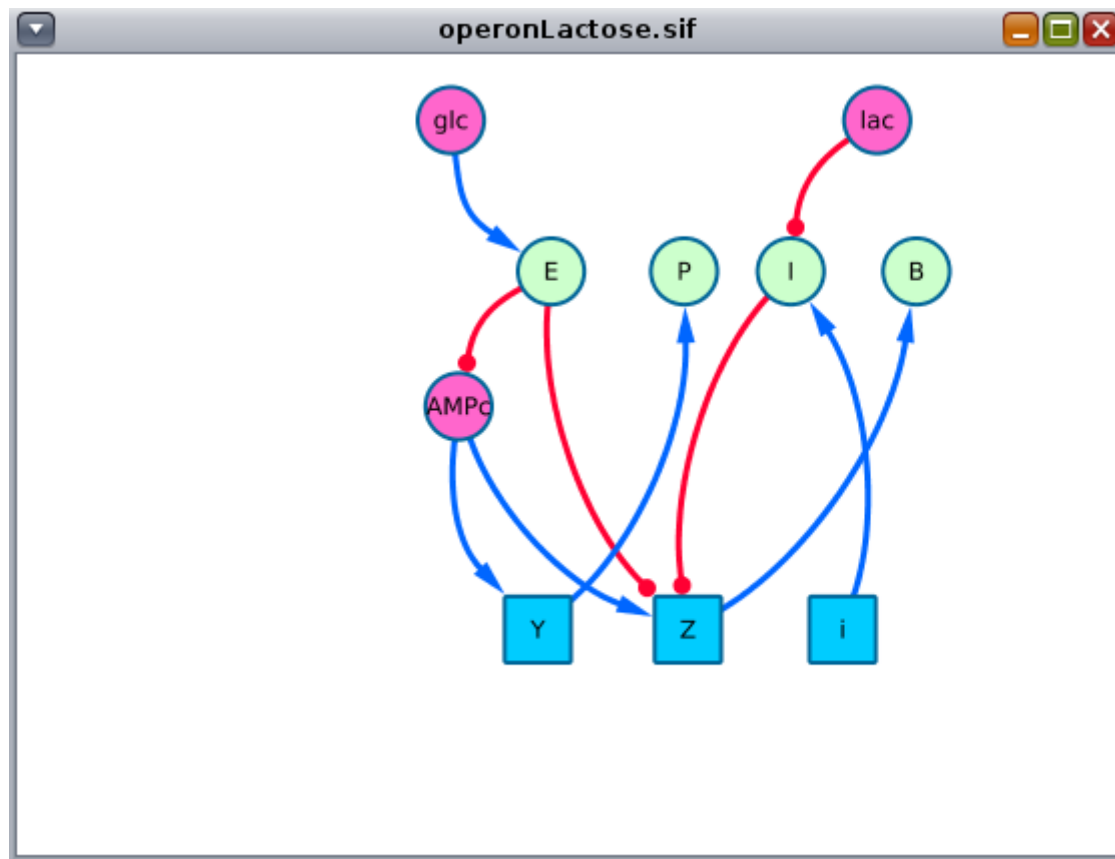
Nous pouvons également choisir de représenter les aretes de façon incurvée en choisissant la propriété Bend et en transformant sa valeur par défaut.



Enfin, changer la couleur des arêtes en fonction de leur nature. Cliquer d'abord sur « Edge color to arrows » puis choisissez la propriété Color.

1.7 Réorganisation des nœuds

Déplacer les nœuds manuellement et avec « Layout→Align And Distribute » pour obtenir une image qui se rapproche le plus de l'original.



Sauvegarder le réseau en fichier image (« File→Export→Network view as graphics »).