

# NARCISSE, une représentation en miroir des synténies conservées.

## Des similarités locales aux segments chromosomiques conservés.

<http://narcisse.toulouse.inra.fr>



Sébastien Letort<sup>1,2</sup>, Emmanuel Courcelle<sup>2</sup>, Olivier Stahl<sup>1</sup>, Jérôme Gouzy<sup>2</sup> et Thomas Faraut<sup>1</sup>

<sup>1</sup> Laboratoire de Génétique Cellulaire INRA  
Thomas.Faraut@toulouse.inra.fr

<sup>2</sup> Laboratoire des Interactions Plantes Micro-organismes, UMR 441-2594 (INRA-CNRS)  
Emmanuel.Courcelle@toulouse.inra.fr  
INRA Toulouse, BP 52627, Chemin de Borde Rouge, Auzeville, 31326 Castanet Tolosan

Avec le nombre croissant de génomes entièrement séquencés, le besoin de pouvoir conduire des analyses comparées sur des séquences complètes se fait de plus en plus pressant. La base de données Narcisse est dédiée à l'étude de la conservation entre les génomes, de l'homologie locale au segment chromosomique conservé ou synténie conservée pour tous les génomes entièrement séquencés organisés en quatre règnes : animaux, plantes, champignons et bactéries.

## Comparaison des génomes : différents niveaux de synténies conservées.

### Comparaison des génomes : les différents niveaux de conservation des synténies.

Au sein d'un même règne, toutes les comparaisons 2-à-2 de génomes sont réalisées aussi bien au niveau nucléaire que protéique. Pour la comparaison nucléaire, un programme nommé glint, développé par deux d'entre nous (Faraut et Courcelle, en préparation), permet une comparaison efficace grâce à un principe d'indexation de génomes. Le programme blast est utilisé pour la comparaison protéique. Les alignements protéiques et nucléiques sont combinés avant de procéder à l'identification des segments conservés.

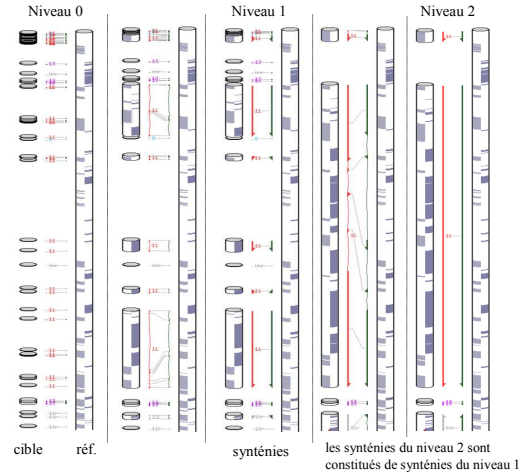
Afin de reconstruire les différents niveaux de synténie conservée un algorithme itératif de clusterisation est appliqué.

A un niveau donné, les segments conservés sont constitués de clusters de segments du niveau inférieur. La construction des clusters repose simplement sur l'identification des composantes connexes d'un graphe. Les segments sont des nœuds qui sont reliés par des arêtes si la distance de Manhattan qui les sépare est inférieure à une distance seuil qui est un paramètre de l'algorithme. Seuls les clusters qui vérifient certaines propriétés de linéarité et de score sont retenus et utilisés comme segments atomiques pour l'étape de clusterisation du niveau supérieur.

Ces 5 représentations à droite, montrent le chaînage permettant de passer d'un niveau à l'autre.

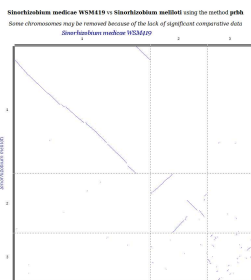
Sur chaque représentation, le cylindre de droite est le chromosome de référence, et à gauche les tranches de chromosome de l'organisme cible. Entre les 2 sont représentés les liens de synténie. Les gènes sont représentés sur les chromosomes.

Le niveau 0 montre une conservation très fine.  
Les premières synténies apparaissent au niveau 1. Entre les deux, on peut voir les segments du niveau 0 qui participent au niveau 1.  
Le niveau 2 montre une vue plus globale avec une grande synténie.



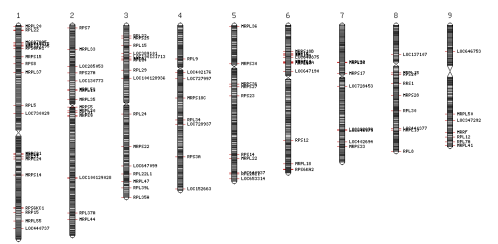
## Le navigateur de génomes comparés Narcisse

Nous proposons différentes façons de visualiser les synténies selon que l'on s'intéresse à une homologie locale ou globale.



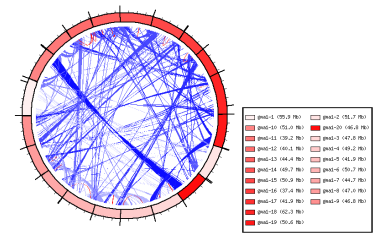
Conservation de l'organisation génomique de deux bactéries fixatrices d'azote du genre Sinorhizobium

### Karyotype for Homo sapiens



Le caryotype permet de visualiser des données personnelles, ici les gènes pour lesquels l'annotation de la protéine contient le mot 'ribosome', sur les chromosomes de 1 à 10 humains.

### Circos using Glycine max V1 as reference



Une représentation circulaire obtenue avec le logiciel Circos [2]. On voit ici les duplications internes chez le soja.

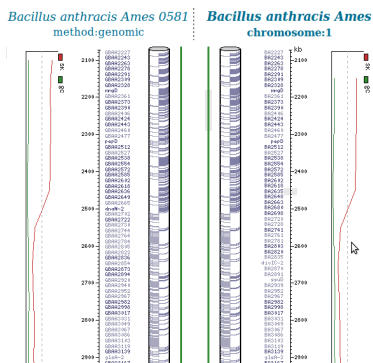
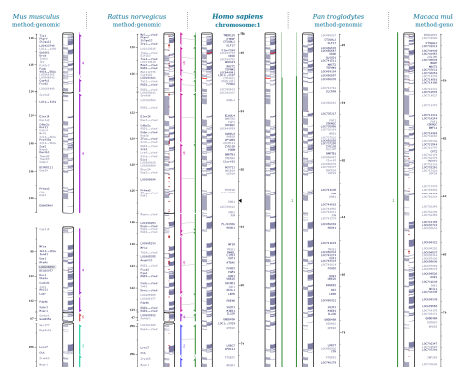


Illustration de l'apport de la représentation de données quantitatives. gc% en vert, biais gc (sk) en rouge. Ici, la chute du biais gc montre la terminaison de la réplication. On peut en particulier remarquer le biais important associé, du brin portant les gènes au voisinage du point de terminaison.



Vue principale représentant, pour une large région du chromosome 1 humain, les conservations de synténies avec les génomes de la souris, du rat, du chimpanzé et du macaque. On peut observer à la fois chez le rat et la souris une inversion de séquence dans la synténie ainsi qu'une rupture de synténie au bas de la figure.

## Technologies

### Mise à disposition des données : les web-services

Afin que les données de Narcisse puissent être exploitées indépendamment du site web, nous avons mis en place des services web permettant d'extraire les informations sur les séquences à l'aide du framework PlayMoby[3]. Certains de ces web-services ont aussi été intégrés à des workflows Remora[4].

### Liens avec d'autres applications

Pour un travail approfondi sur les légumineuses, nous proposons des liens vers des applications spécifiques. (voir le poster LEGoo).

### Technologies informatiques

Notre navigateur de génomes est utilisable via Firefox et Internet Explorer. Afin d'améliorer l'interactivité de l'interface, nous faisons appel à des requêtes AJAX en utilisant les bibliothèques prototype[5] et YUI[6]. Narcisse utilise les langages c++, Perl (objet), javascript et xhtml.

## Références

- [1] Courcelle E, Beausse Y, Letort S, Stahl O, Fremez R, Ngom-Bru C, Gouzy J, Faraut T (2008). Narcisse : a mirror view of conserved synteny. Nucleic Acids Res., 36 :D485-90.
- [2] Krzywinski M. <http://jmkweb.bcgsc.ca/circos/>.
- [3] Playmoby : <http://ijpm-bioinfo.toulouse.inra.fr/biomoby/playmoby/>
- [4] Carere S, Gouzy J. (2006) REMORA : a pilot in the ocean of BioMoby web-services. Bioinformatics 22 :900-1.
- [5] Prototype : <http://www.prototypejs.org/>
- [6] YUI : <http://developer.yahoo.com/yui/>

## Remerciements

Nous remercions le réseau d'excellence Rex eadgène pour son soutien financier.

Le projet Narcisse a été financé grâce aux projets ANR :  
GPLA06026G ANR Genoplante  
ArcAnge ANR Genanimal project

Le projet Narcisse a bénéficié des ressources informatiques du  
Géopôle Toulouse Midi-Pyrénées

