

MABioVis

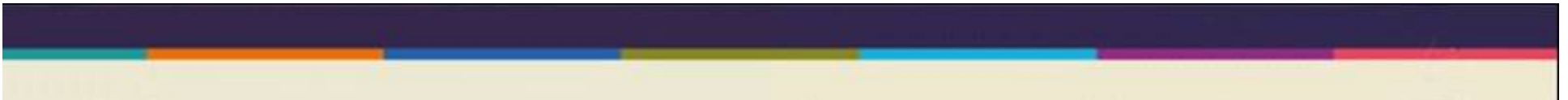
Modèles et Algorithmes pour la Bio-informatique et la Visualisation



Visite ENS Cachan 5 janvier 2011

MABioVis

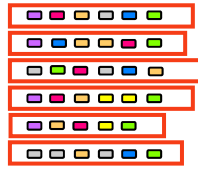
- GUY MELANÇON (PR UFR Maths Info / EPI GRAVITE)
(là, maintenant) - MABioVis
- DAVID SHERMAN DR INRIA (EPI MAGNOME)
Mercredi 17h – Génomique
- DAVID AUBER (MCF UFR Maths Info / EPI GRAVITE)
Jeudi 10h – Visual Analytics
- BRUNO PINAUD (MCF UFR Maths Info / EPI
GRAVITE) Jeudi 11h-13h « Réseaux biochimiques
évolutifs et visualisations de graphes » (discussion)



MABioVis

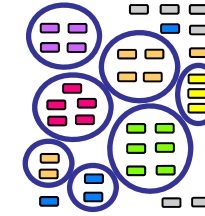
- La bio-informatique comme domaine d'application cible
- Objectifs complémentaires
 - Travailler au plus près des données biologiques, concevoir et mettre en œuvre des modèles
 - Apporter des objets/modèles combinatoires à des fins algorithmique, d'analyse et de modélisation
 - Développer des méthodes offrant ultimement à l'expert un laboratoire d'exploration et d'analyse des données



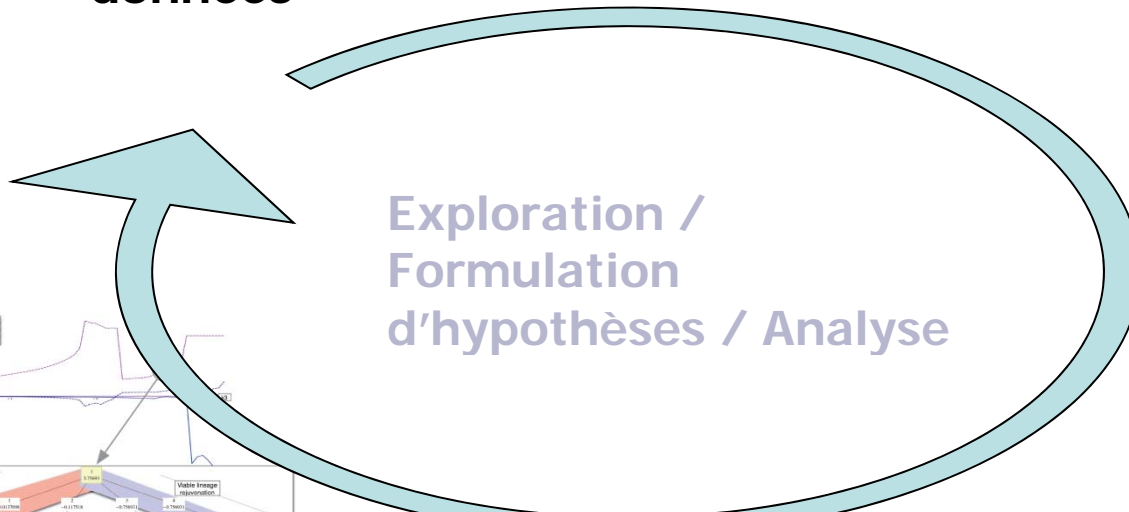


Production de données

Données biologiques



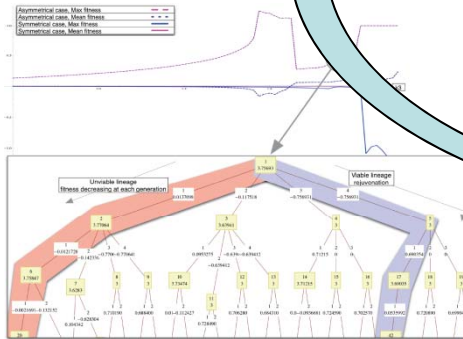
Annotation primaire



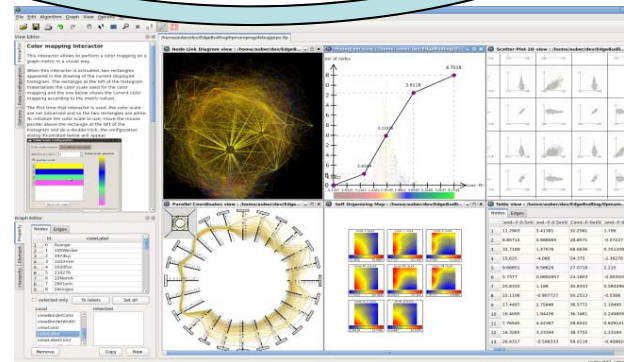
**Exploration /
Formulation
d'hypothèses / Analyse**

**Analyse
et algorithmes**

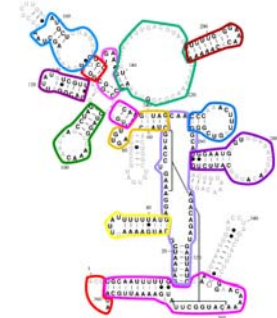
**Objets/modèles
(combinatoires)**



Elaboration de modèles

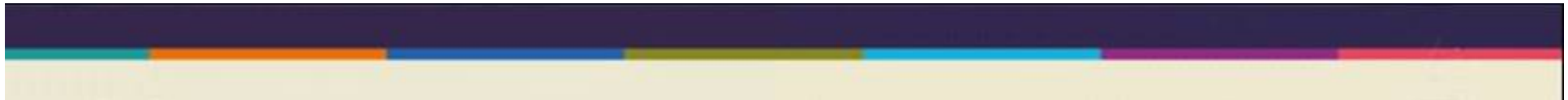


**Intégration
et visualisation**



MABioVis – nos fondamentaux

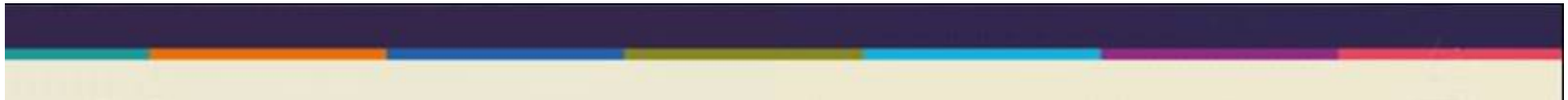
- Apport des fondamentaux informatiques
 - Méthodes formelles
 - Théorie des graphes, mathématiques combinatoires
 - Optimisation combinatoire
- Abstraire la méthodologie des objets modélisés
 - Structures et algorithmes des mathématiques discrètes
 - Permutations et motifs interdits
 - Arborescences multi-échelle
 - Graphes planaires, petit monde, invariant d'échelle



La bio-informatique – domaine fédérateur et catalyseur

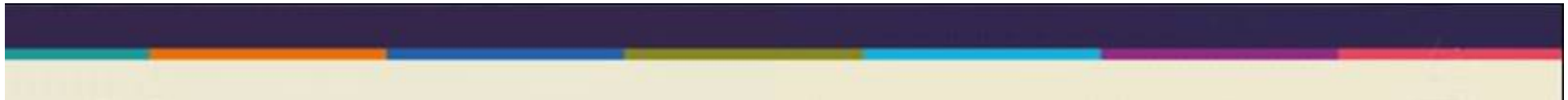
- Comparaison de génomes / structure, évolution
 - Deux à deux: combinatoire des structures
 - Multiple: optimisation combinatoire
- (Re) construction de réseaux biologiques
- Modélisation multi- échelle
 - Représentation des connaissances
- Intégration de données

- Approche exploratoire
- Confirmation d'hypothèses



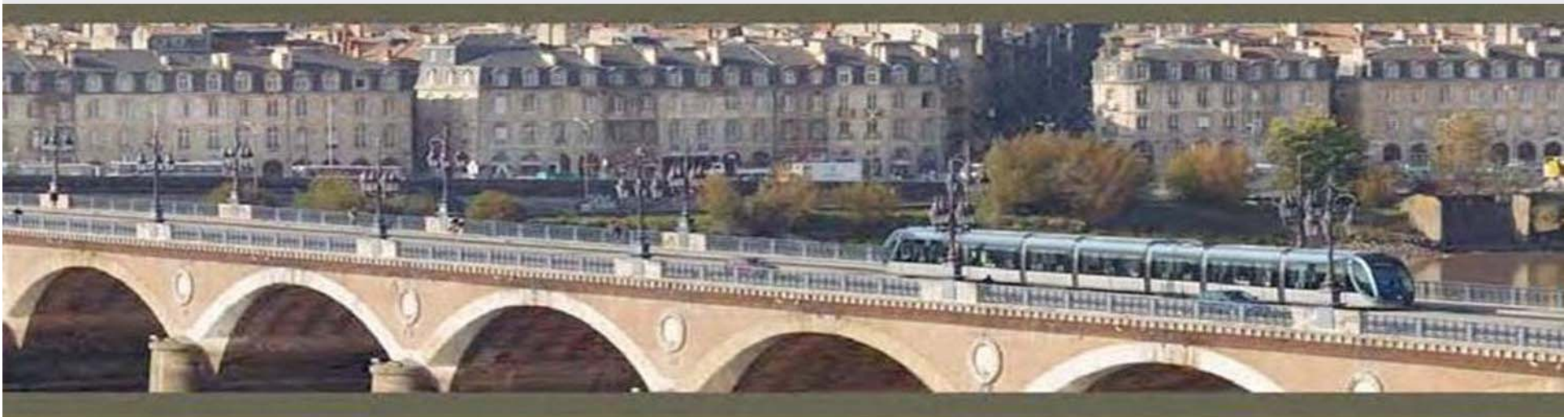
MABioVis –méthodologie(s)

- Calcul de similarités (sur des objets ou sur des groupes d'objets)
- Réduction de dimension, clustering
- Autosimilarité, compression, échantillonnage, passage à l'échelle
- Analyse des séquences et du texte (structures linéaires), des structures arborescentes
- Objets/mesures combinatoires
 - Formulation des algorithmes
 - Analyse des algorithmes – Complexité, convergence ou passage à l'échelle des algorithmes

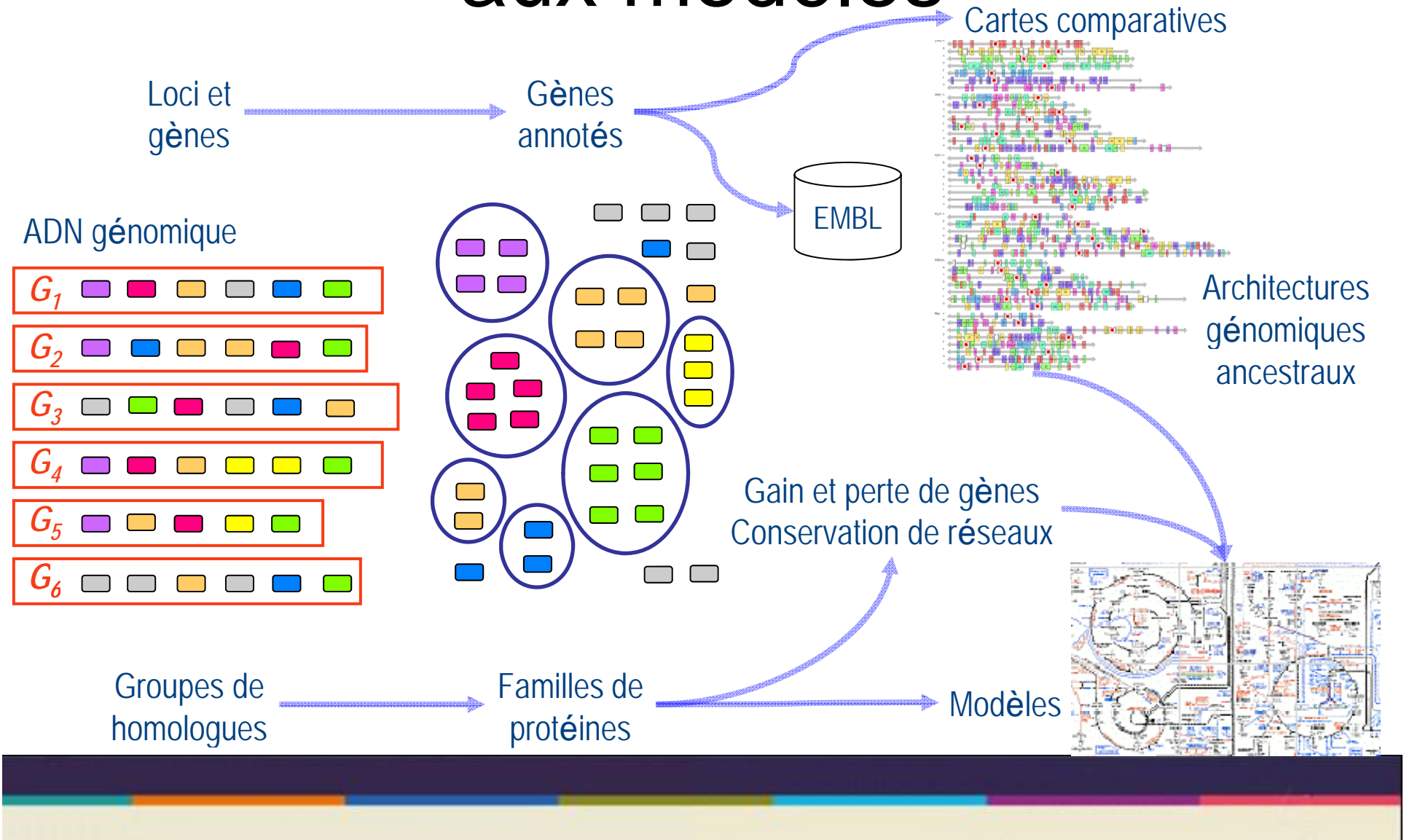


Bilan Scientifique

Quelques résultats



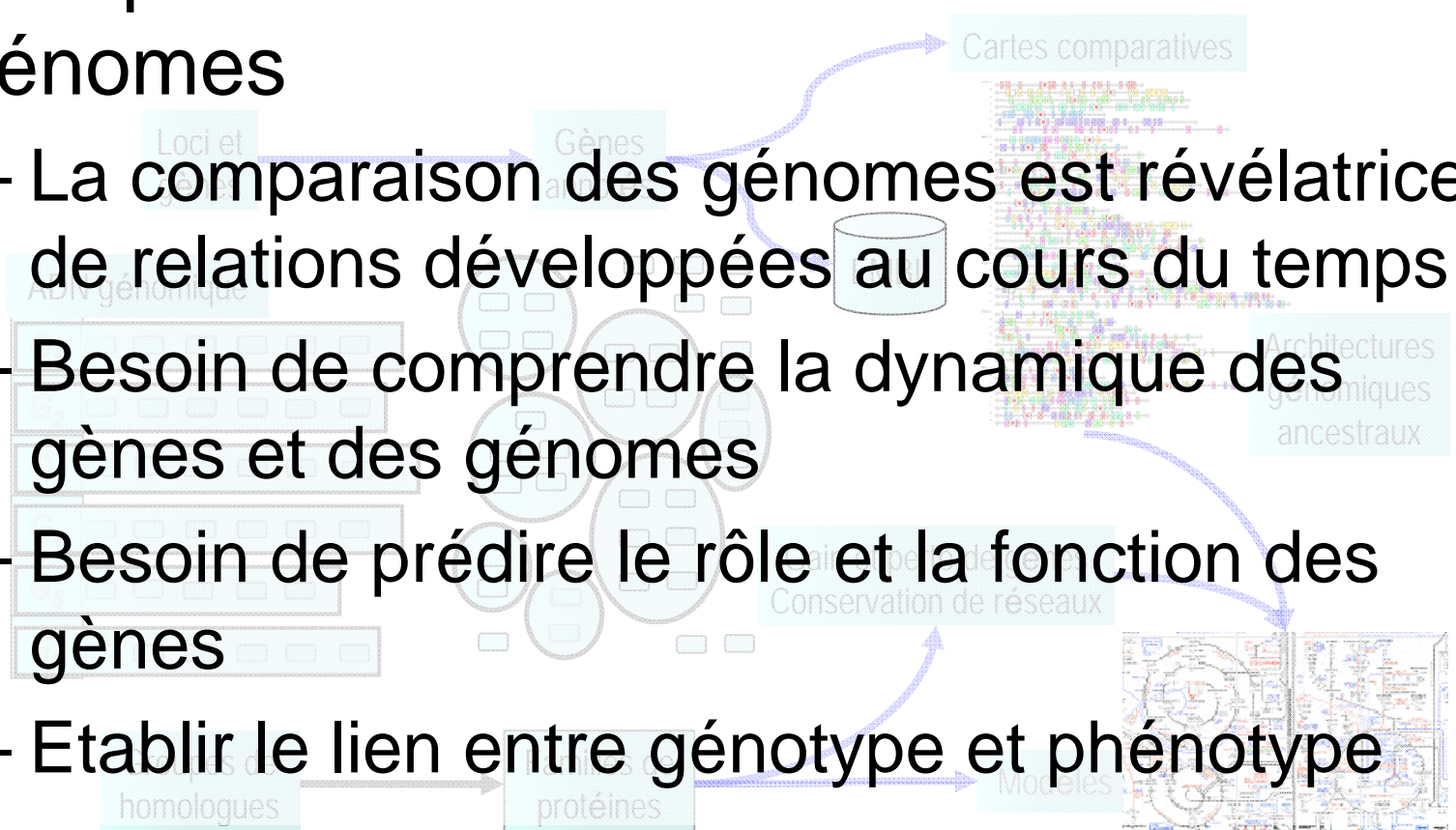
De la comparaison de génomes aux modèles



De la comparaison de génomes aux modèles

- Comprendre l'évolution et la fonction des génomes

- La comparaison des génomes est révélatrice de relations développées au cours du temps
- Besoin de comprendre la dynamique des gènes et des génomes
- Besoin de prédire le rôle et la fonction des gènes
- Etablir le lien entre génotype et phénotype



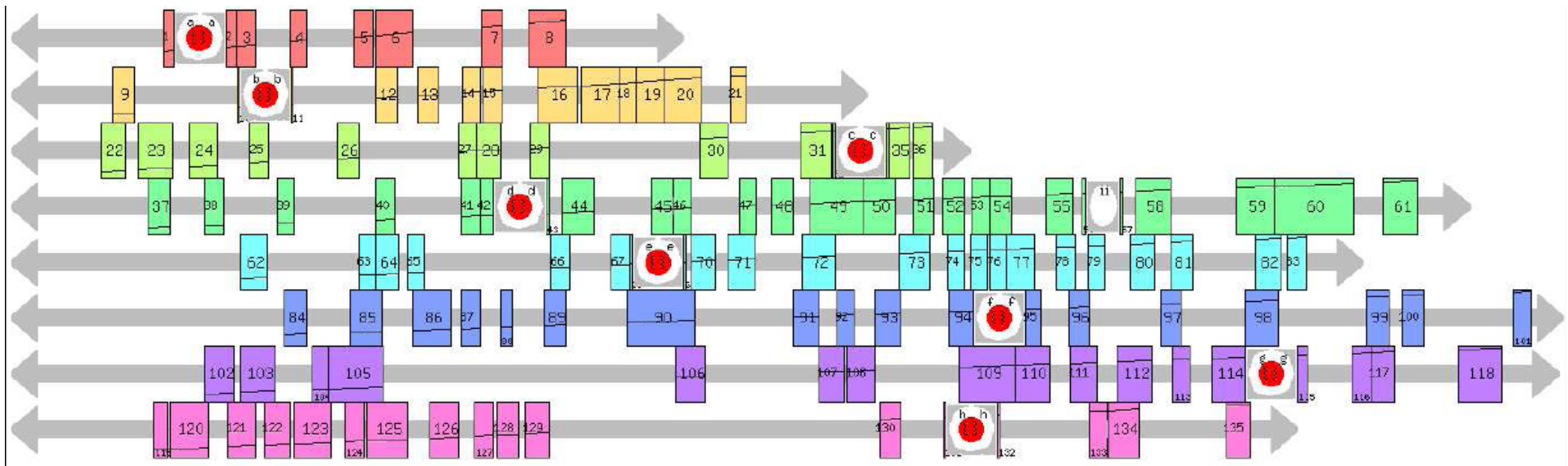
Comparaison de génomes

- Clustering par consensus
 - Eviter les écueils d'un choix particulier
 - Rechercher la « robustesse » de la partition
 - Définir une heuristique qui se calcule en temps raisonnable

Consensus clustering: étant données des partitions P_1, \dots, P_k , trouver une partition consensus P minimisant $\sum d(P, P_k)$ où d est une distance.

Comparaison de génomes

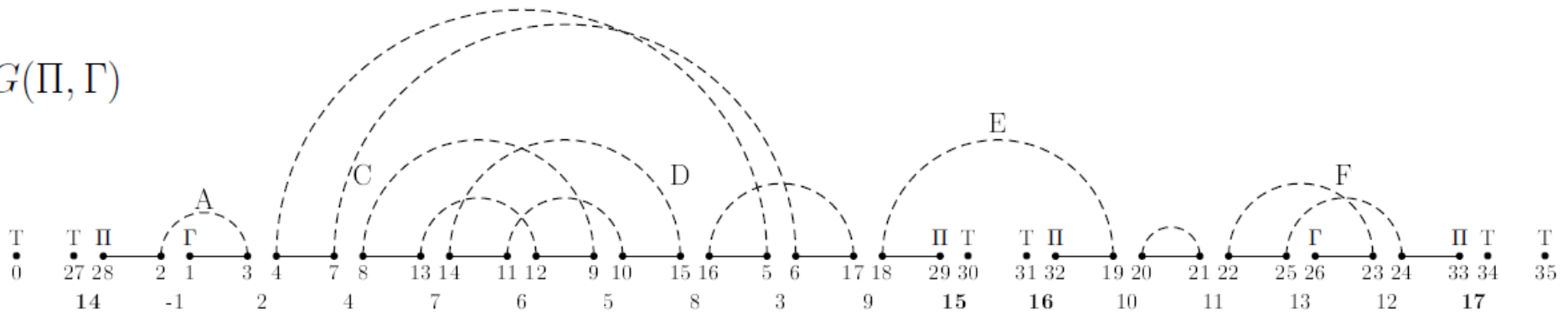
- Comparaison et réarrangements de plusieurs génomes
 - modélisés par des permutations signées



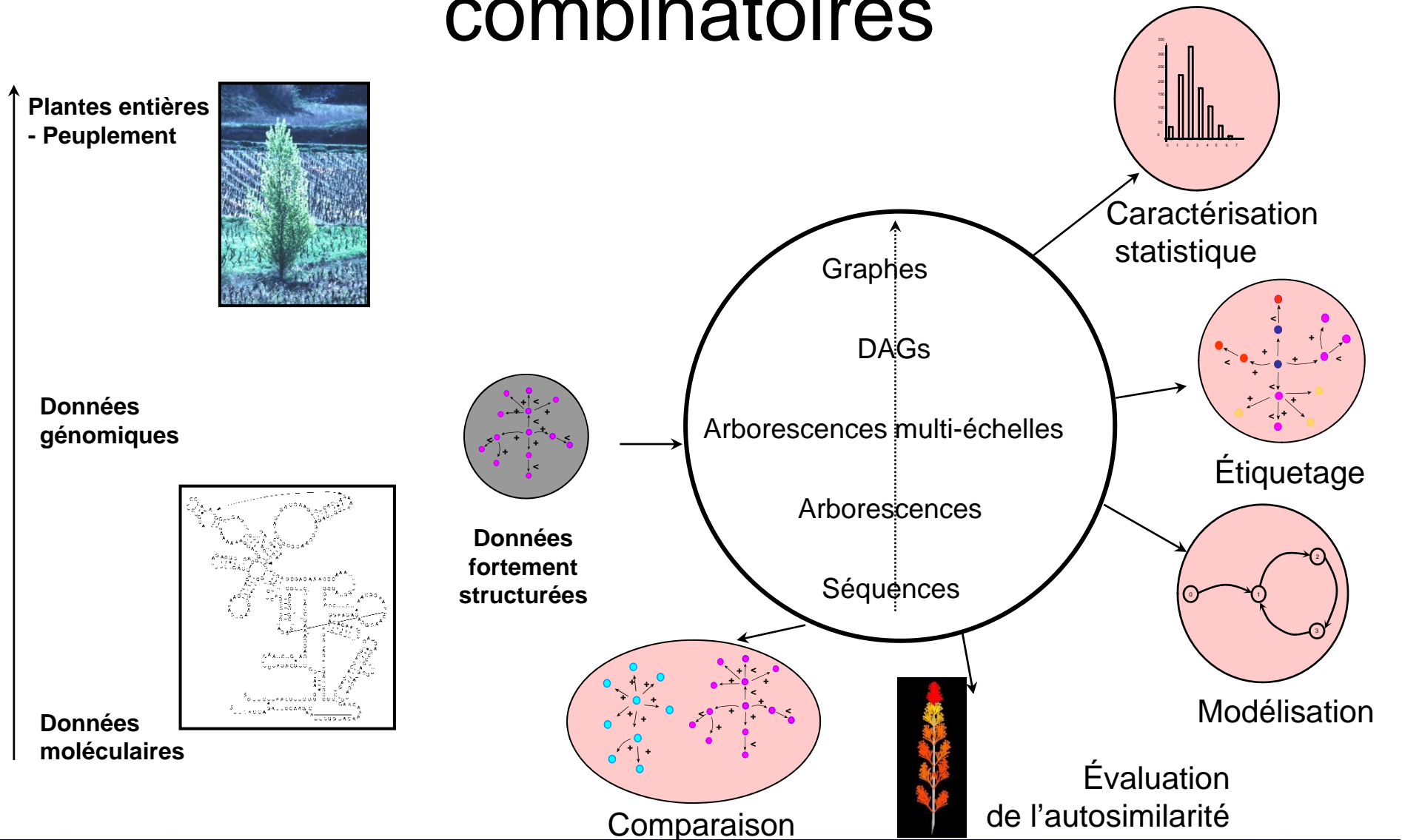
Comparaison de génomes

- Comparaison et réarrangements de plusieurs génomes
 - Génome médian : à distance minimal de tous les génomes considérés ()
 - Adjacence préservées / perdues encodées par un graphe (breakpoint graph)

$G(\Pi, \Gamma)$

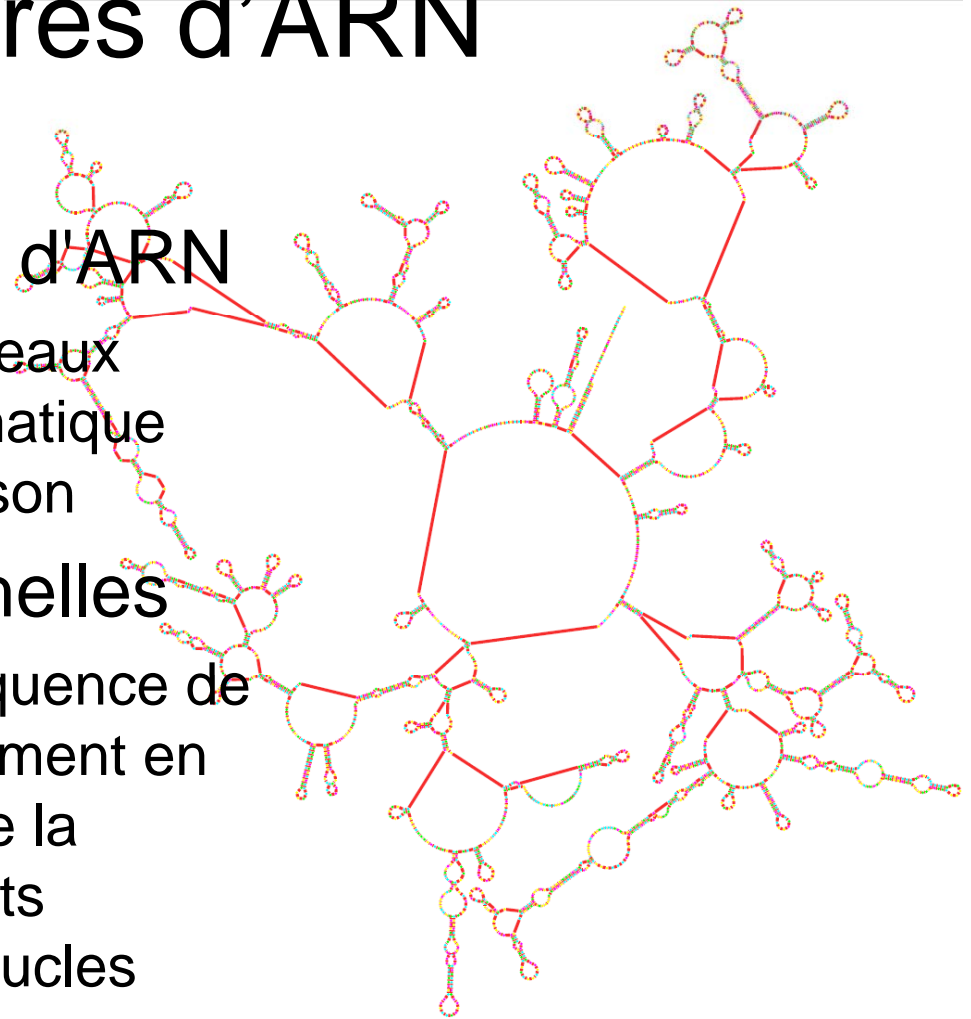


Analyse de structures biologiques combinatoires



Comparaison de Structures Secondaires d'ARN

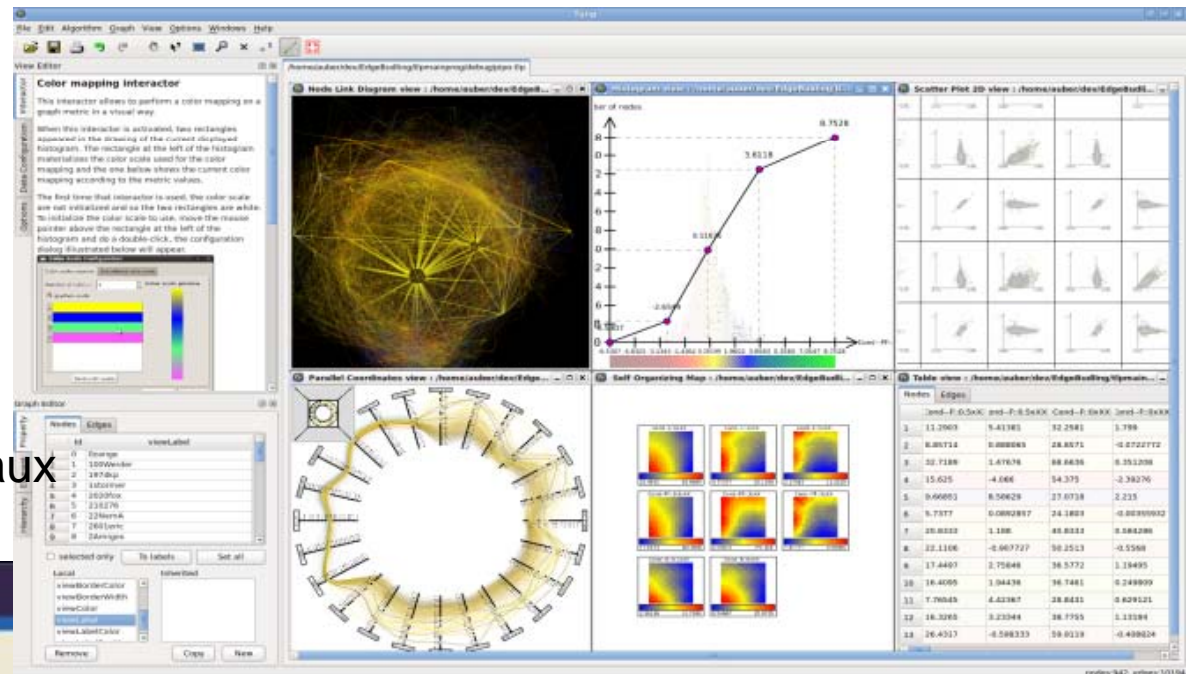
- Bases de données de structures secondaires d'ARN
 - Important besoin de nouveaux outils de traitement automatique d'analyse et de comparaison
- Comparaison multi-échelles
 - Prise en compte de la séquence de nucléotides, de son repliement en structure arborescente, de la décomposition en éléments structuraux (réseau de boucles multiples)



Visualisation d'information

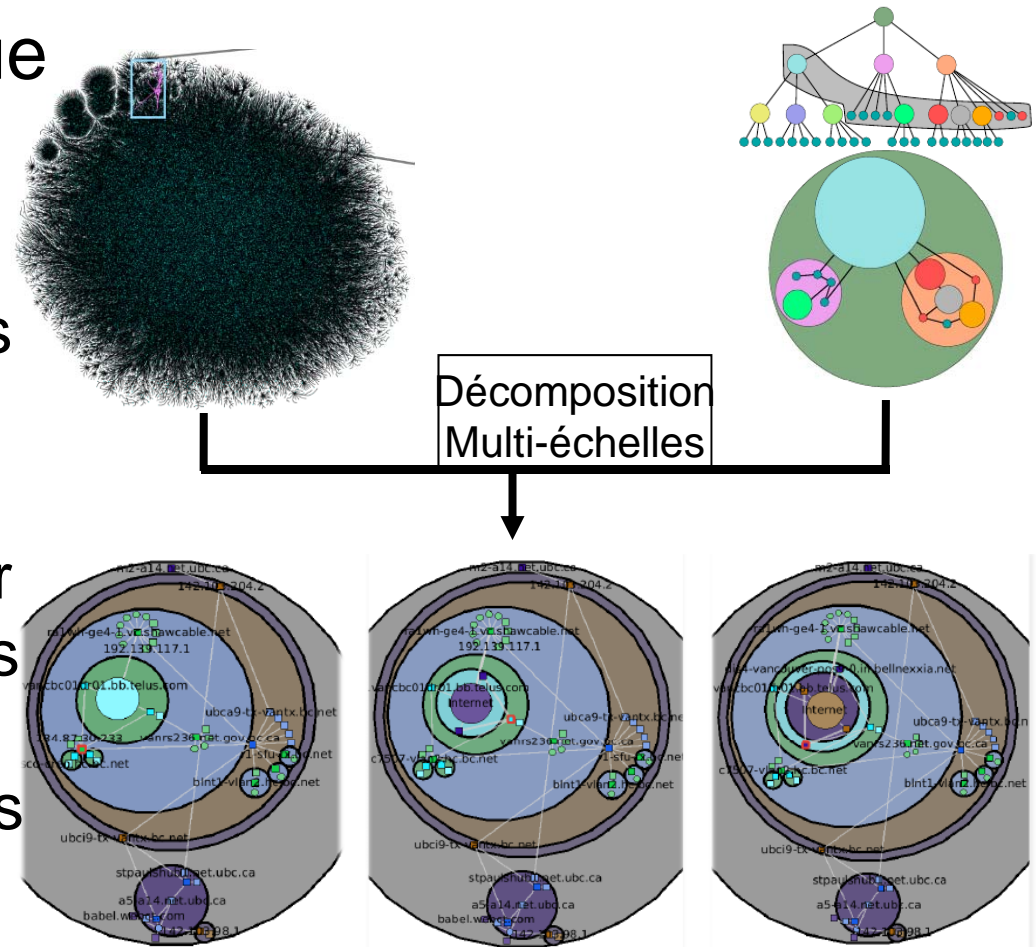
- La visualisation comme stratégie d'exploration et de formulation d'hypothèses
 - Graphes petits mondes / invariant d'échelle
 - Approches (semi) supervisées

Algorithmes de dessin
Exploitation des indices structuraux
Interactions spécifiques



Visualisation multi-niveaux

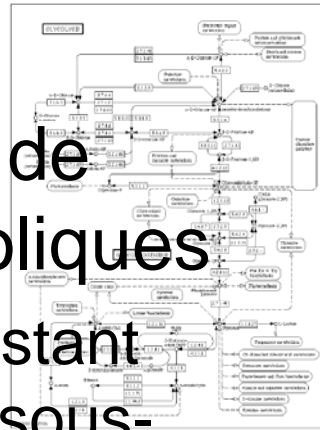
- Approche topologique pour l'exploration de grands graphes
 - Algorithmes existants peu efficace sur les très grands graphes
 - Besoin de rechercher automatiquement des sous-structures topologiques dans les données



Visualisation multi-niveaux

- Représentation de réseaux métaboliques

- Algorithmes existant travaille sur un sous-ensemble du réseau
- Nouvel algorithme permettant d'éviter la duplication de sommets en conservant au mieux la représentation des chemins métaboliques



Analyse topologique
Dessin planaire

