Annotation des génomes complets

Seminaire IN'Tech
bioinformatics : from genomics and post-genomic data to biological knowledge

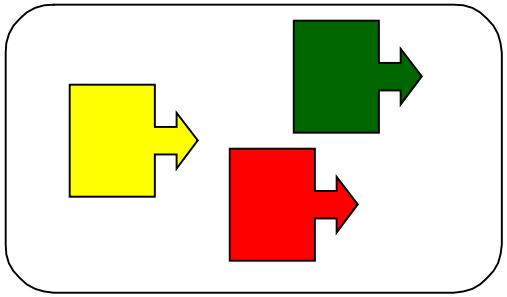
Yves Vandenbrouck

GENOME express

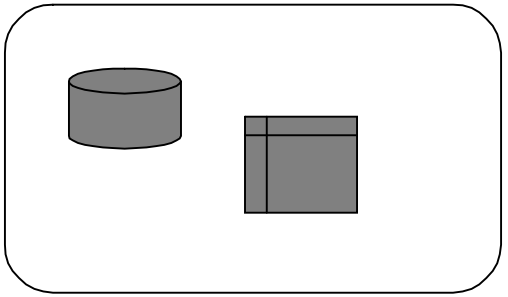
23 octobre 2003, Lyon

Environnements d'annotation (1)

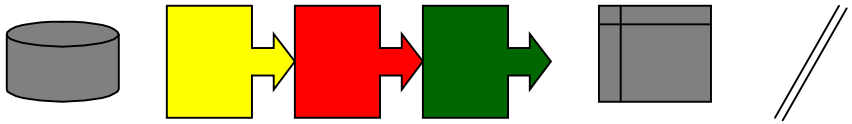
outils



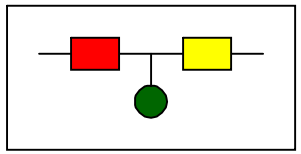
données



- « pipe-line »



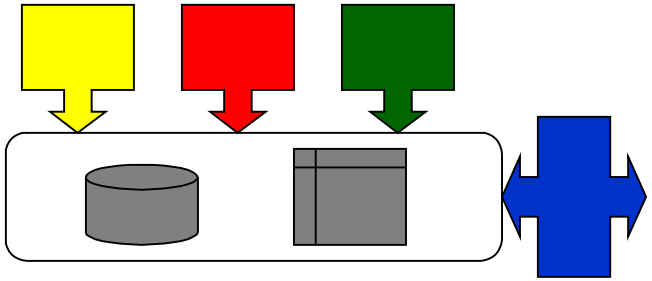
Interface



- gros débits
- programmation
- interfaces « passives » (consultation)

ex: Pedant, GeneQuiz, BioFacet

- annotateurs automatiques



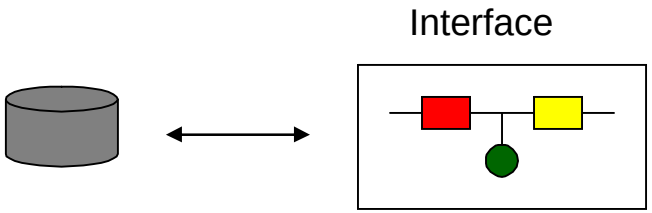
synthèse
(règles)

Magpie (Gaasterland et al.)

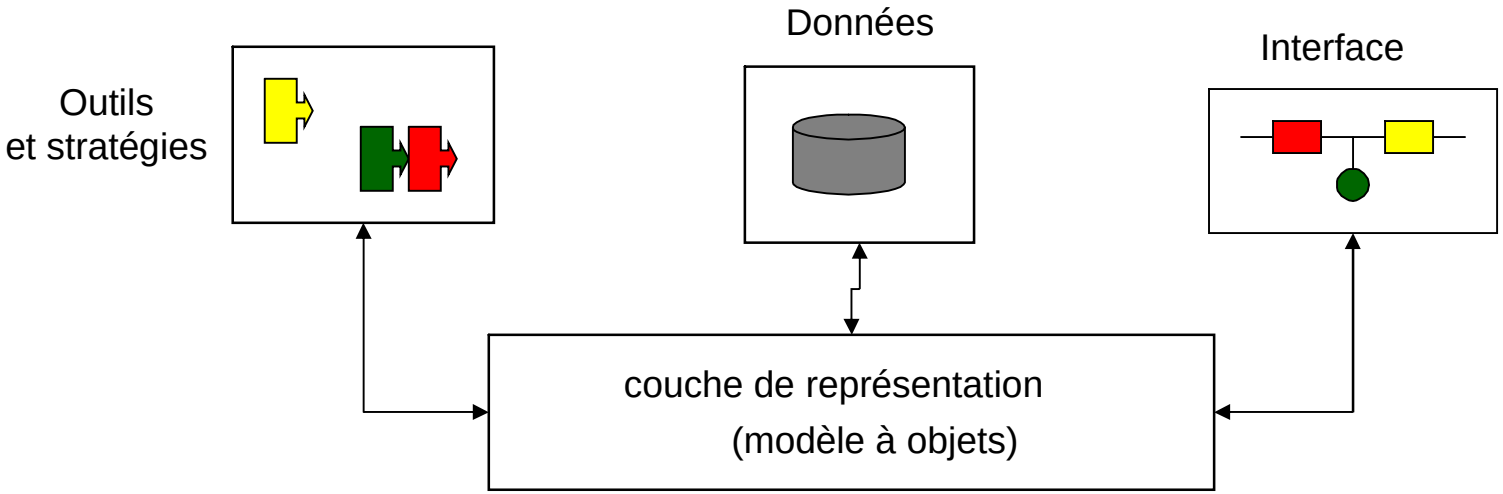
Environnements d'annotation (2)

- éditeurs d 'annotations

  Genotator (Harris et al.)
  Artemis (Rutherford et al.)

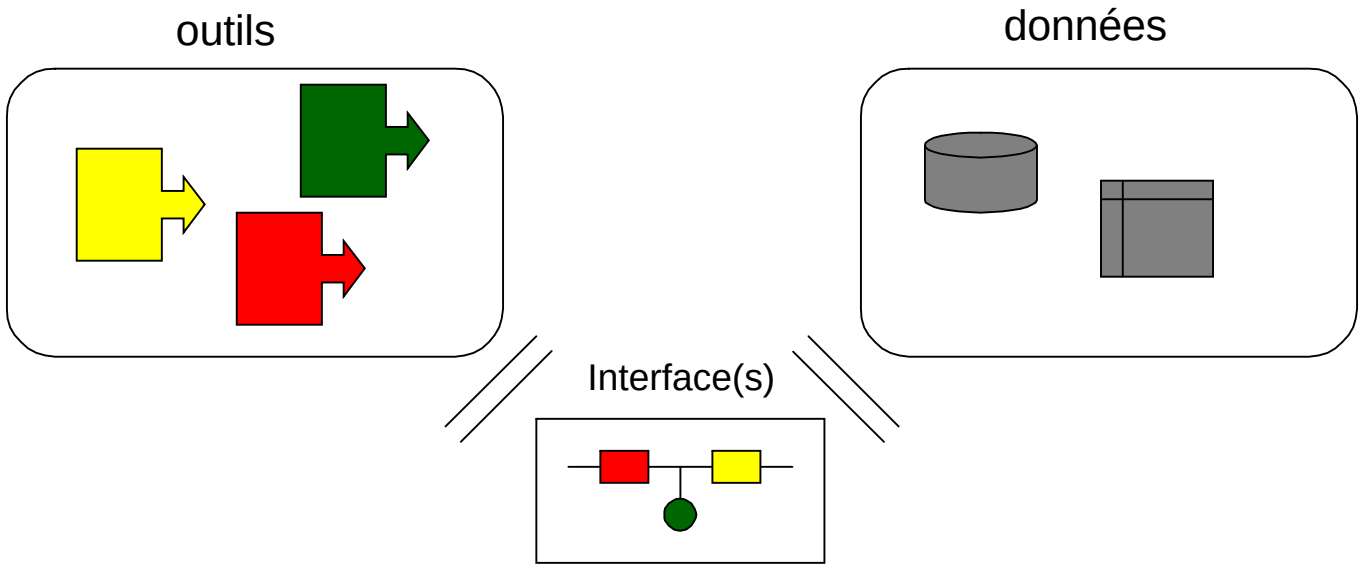


- environnements interactifs



Imagene (Médigue et al. 99)

Environnement d'annotation : les objectifs

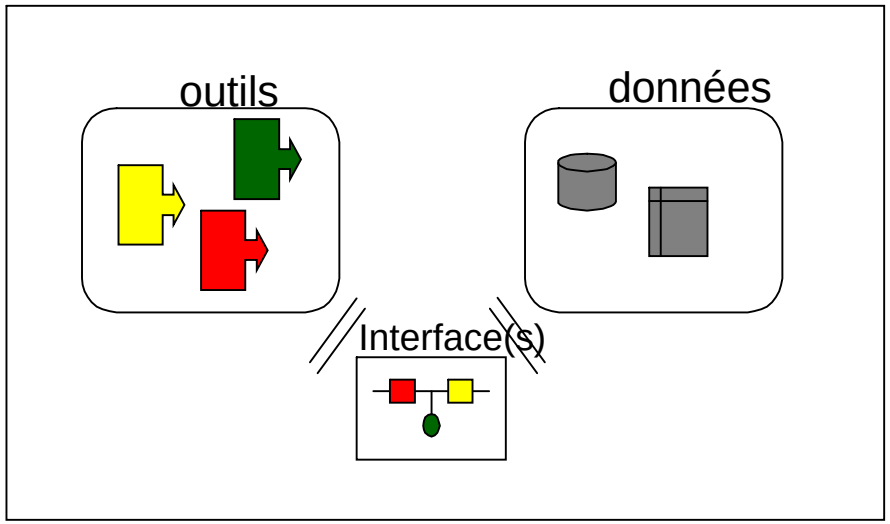


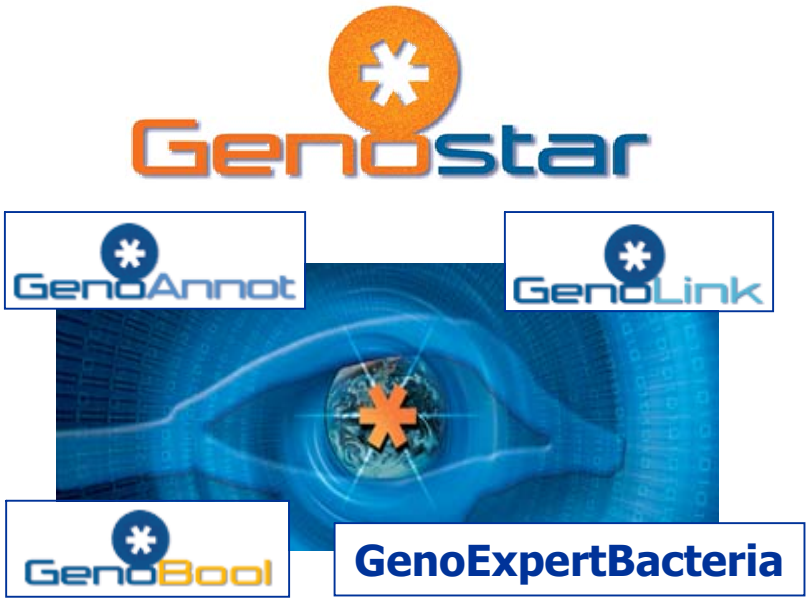
1. Confronter différentes sources de données : Hétérogènes, distribuées,

2. Appliquer des méthodes d'analyse dédiées : Stratégies

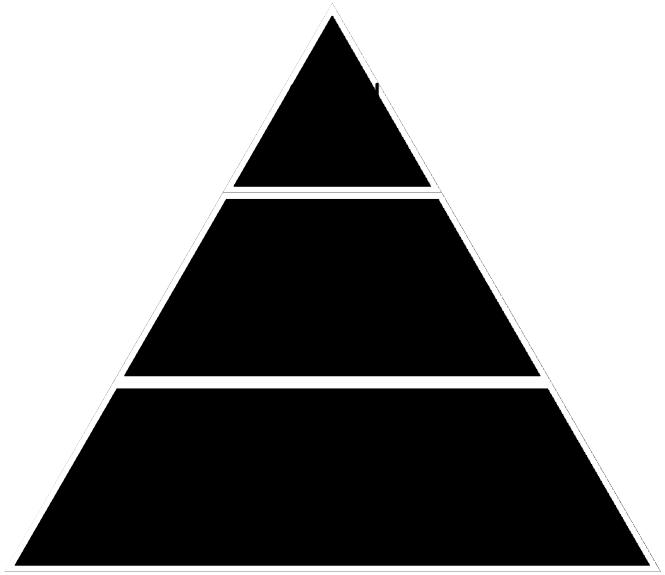
3. Favoriser le travail de l'expert-annotateur : Interfaces utilisateur (contrôles des paramètres, visualisation des résultats, édition des annotations)

L’environnement GenoStar



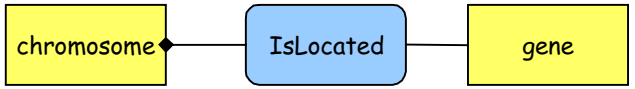


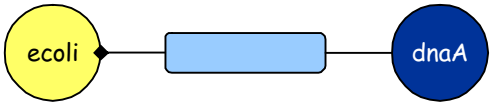
GenoStar : Système de représentation des connaissances



Entity / relation
Classes/Associations

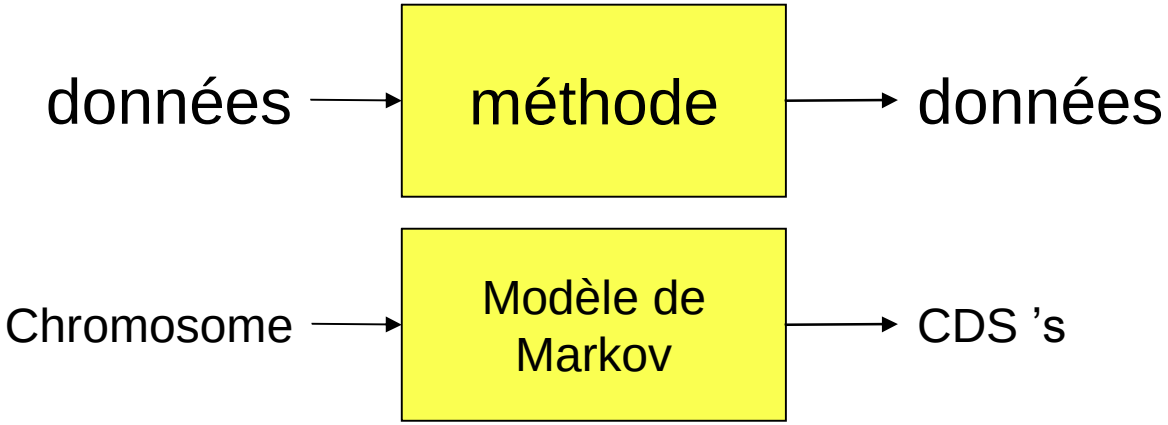
UML





Object Oriented Representation System

GenoStar : Méthodes (1)



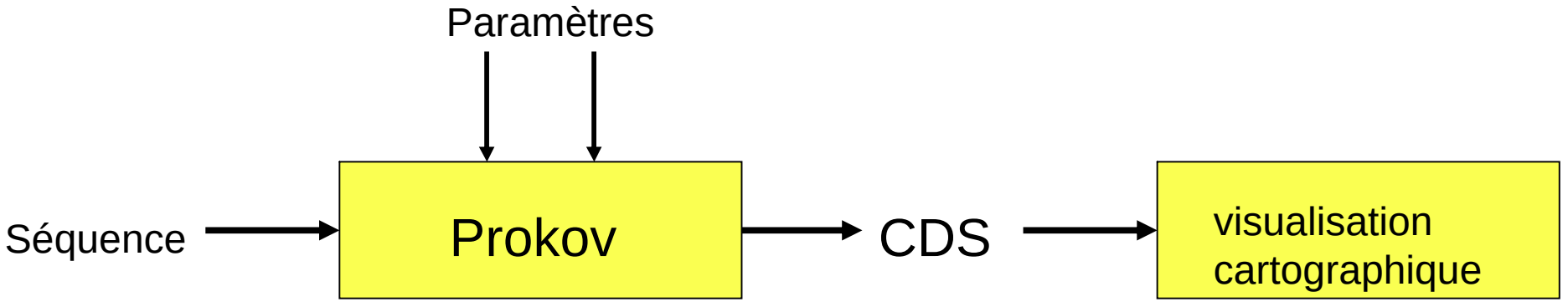
- représenter et organiser les méthodes comme des entités

  Modélisation Connaissance méthodologique

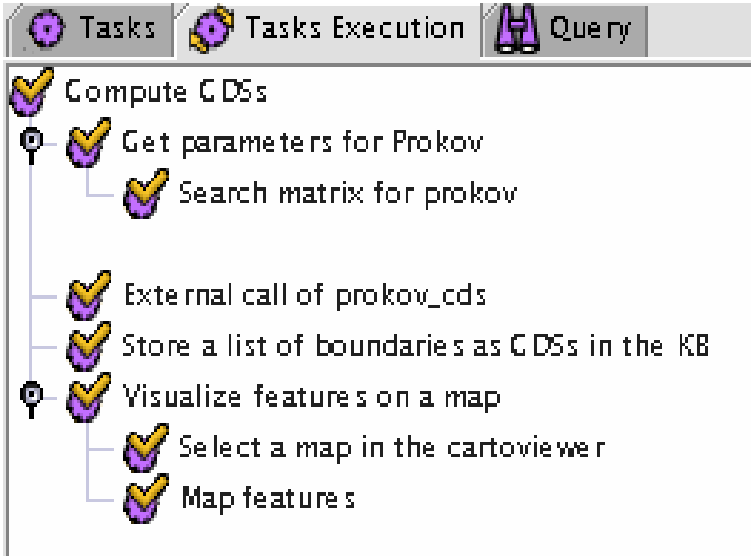
- contrôler l'exécution (quelle méthode pour quelle donnée)

GenoStar : Méthodes (2)

• Des tâches paramétrables enchaînant ces méthodes sur les données

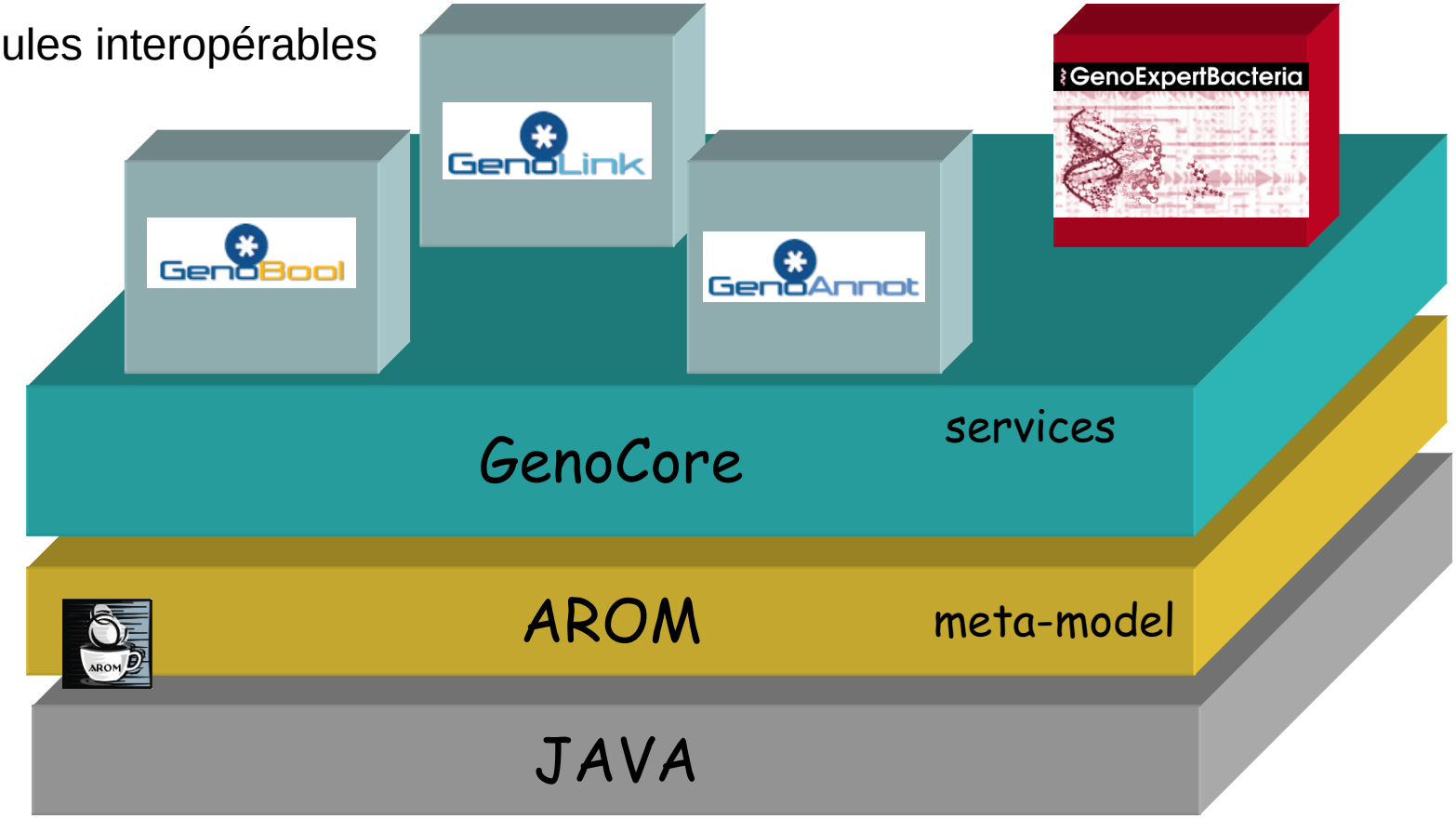


Décomposition logique



GenoStar : Architecture

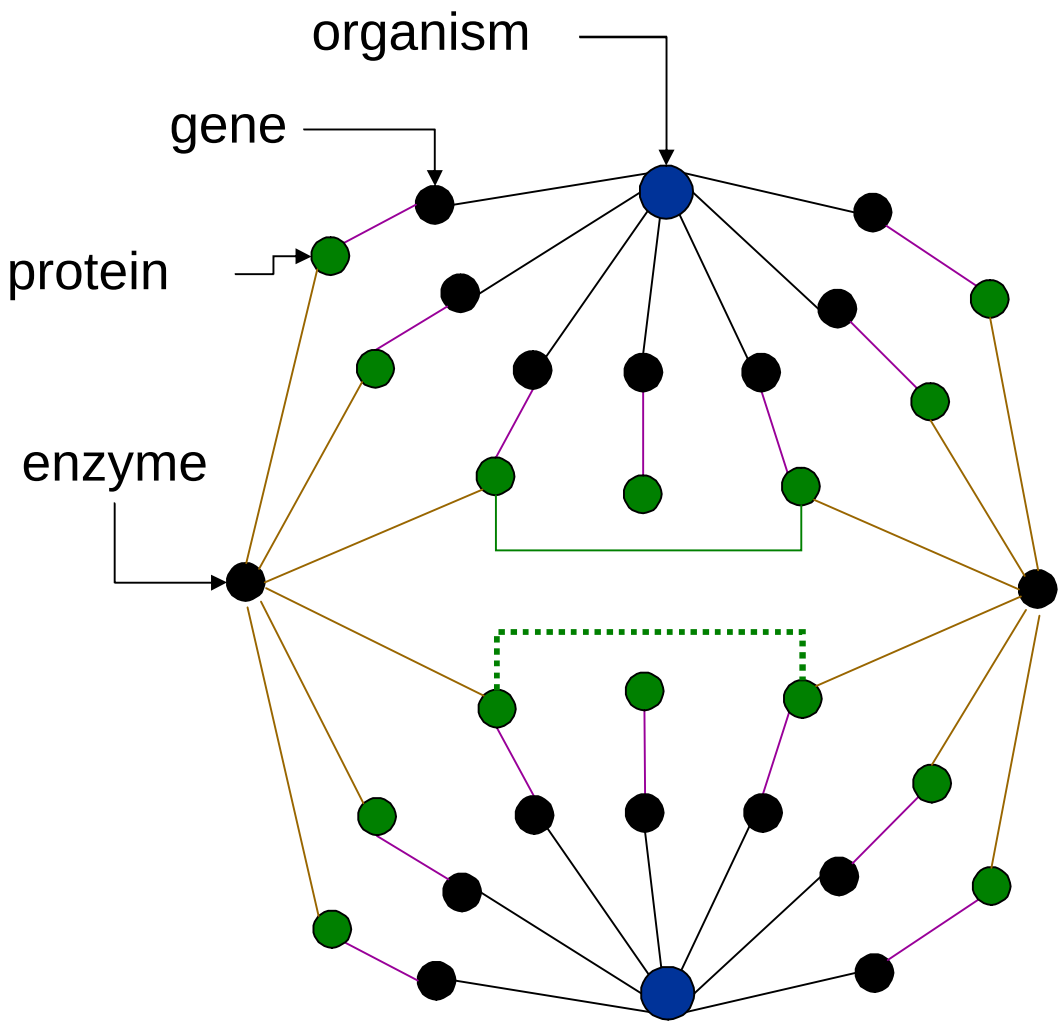
Des modules interopérables



autour d'un noyau qui assure :

- la gestion des données et des connaissances
  - l'enchaînement de l'exécution des méthodes d'analyse

GenoStar : Environnement de génomique exploratoire



Annotation de séquences

- **Identifier les objets**

GenoAnnot

Annotation fonctionnelle

- **Explorer le voisinage**

GenoLink

GenoExpertBacteria

Fouille de données

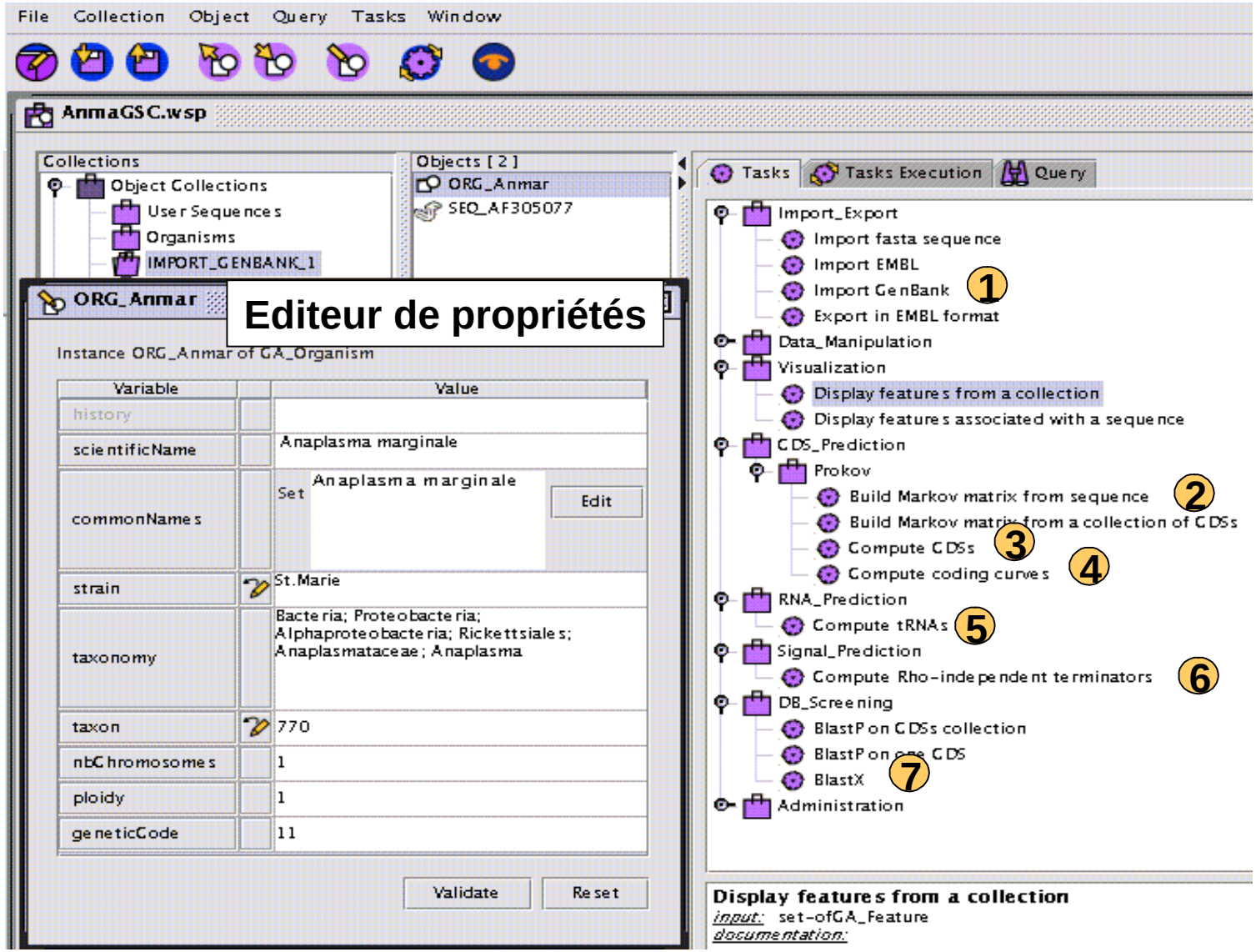
- **Découvrir de nouveaux objets / relations**

GenoBool

GenoAnnot : annotation séquence brute (1)

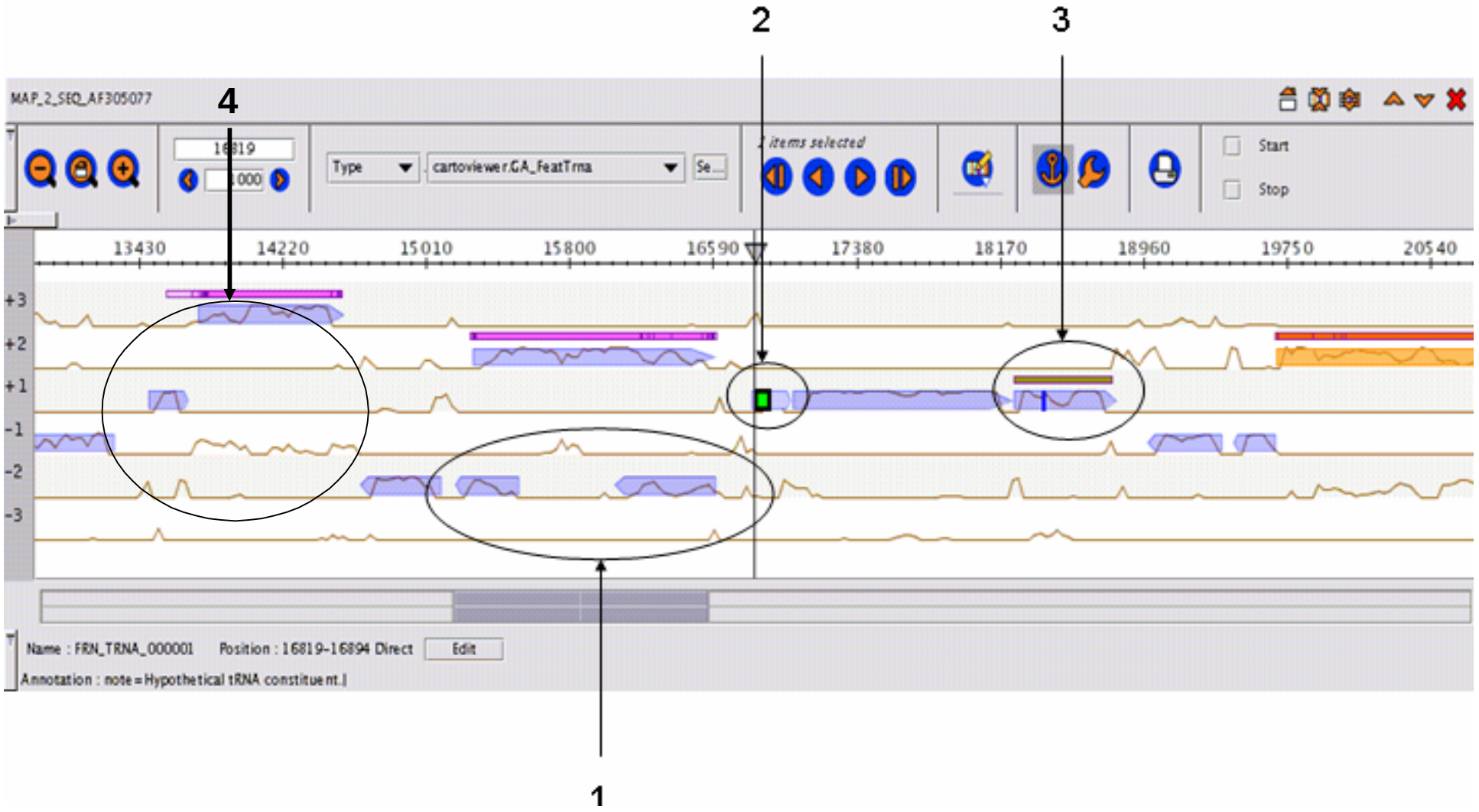
- Gestion des données et des taches :

  Contig ~ 45 kb (AF305077, GenBank) *Anaplasma marginale*



GenoAnnot : annotation séquence brute (2)

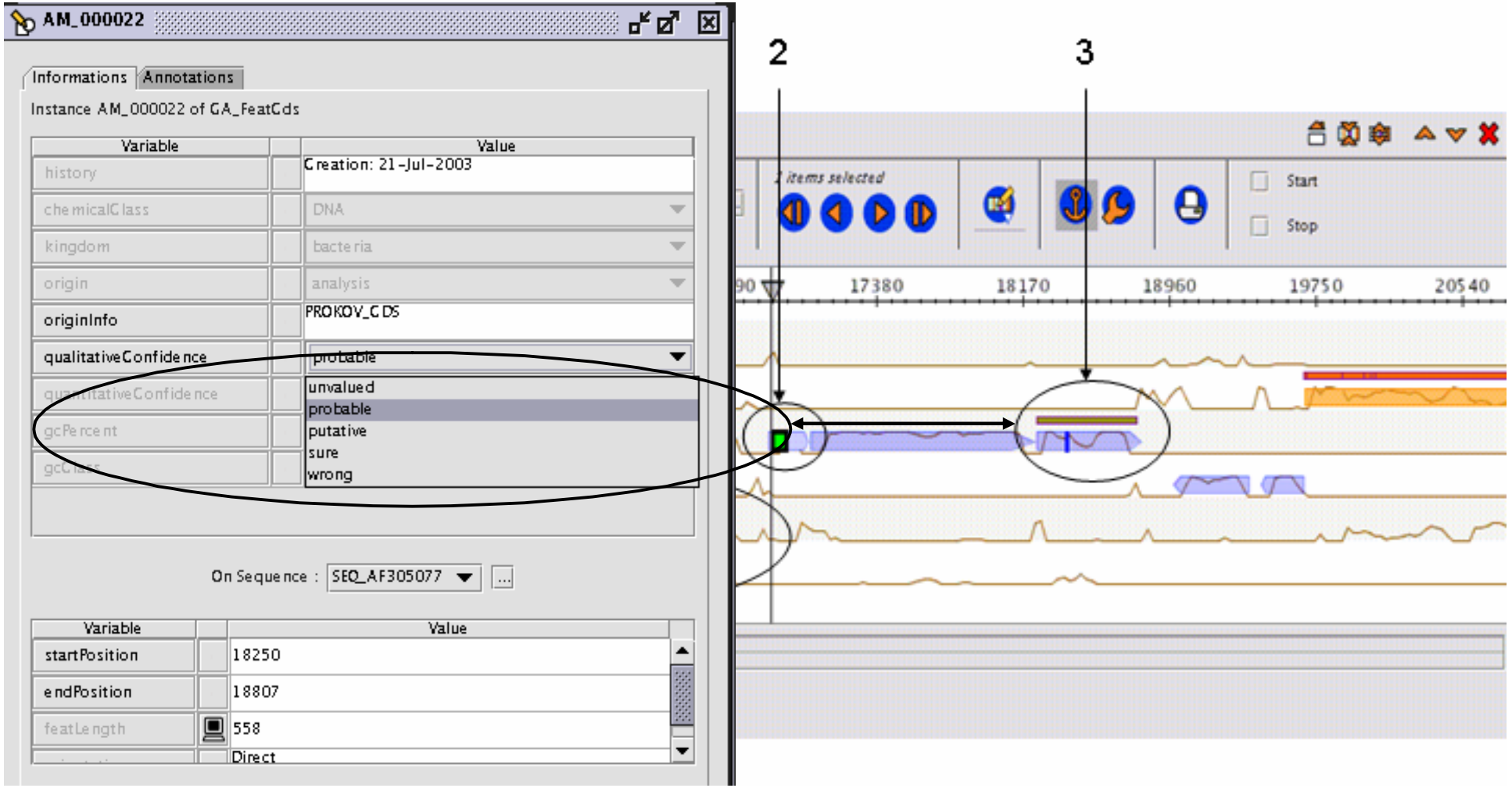
- Interface cartographique : visualisation des résultats des méthodes



GenoAnnot : annotation séquence brute (3)

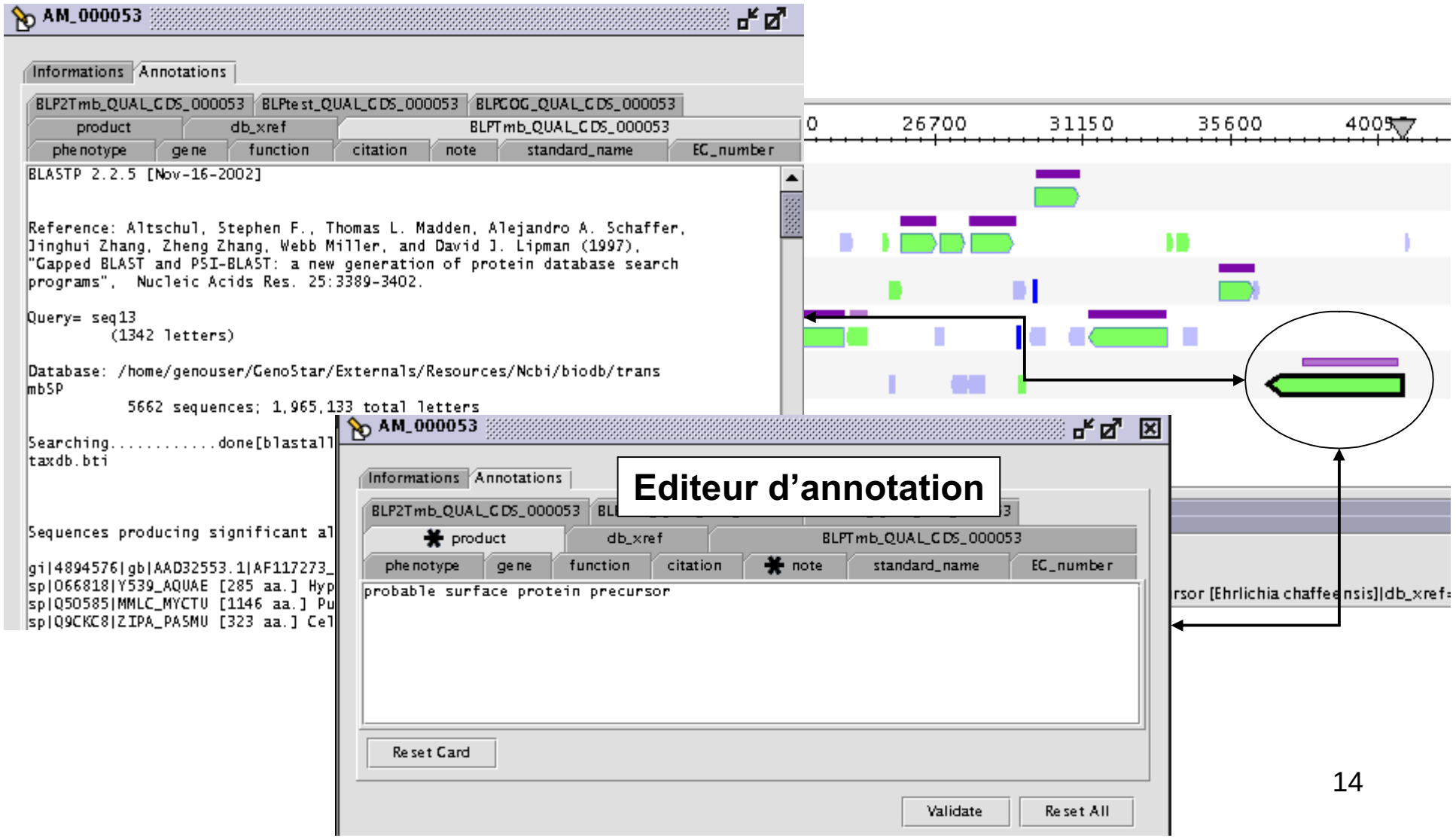
- Interface cartographique : visualisation des résultats des méthodes

- Editeur de propriétés : statut des features



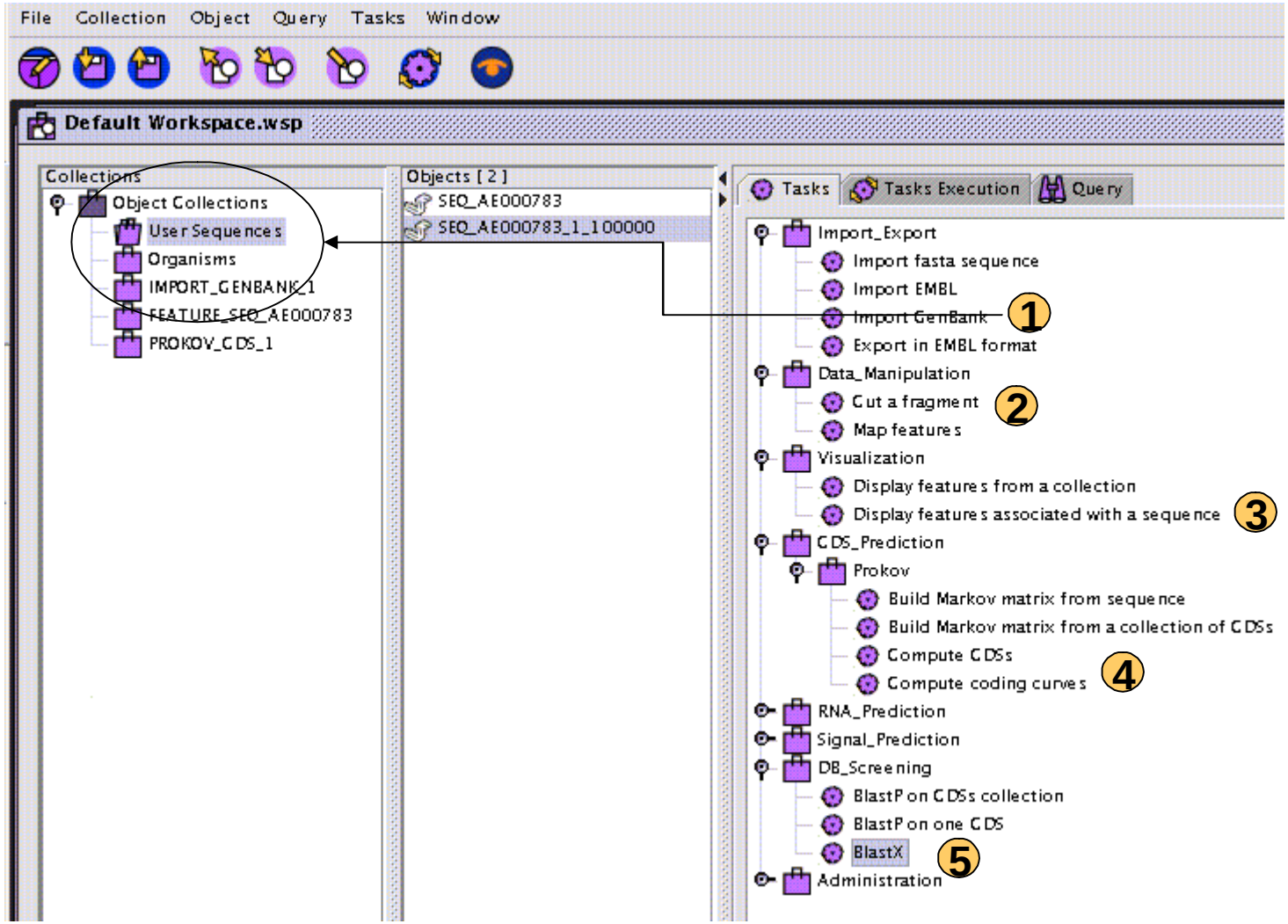
GenoAnnot : annotation séquence brute (2)

Utilisation de blastp et assignation fonctionnelle primaire



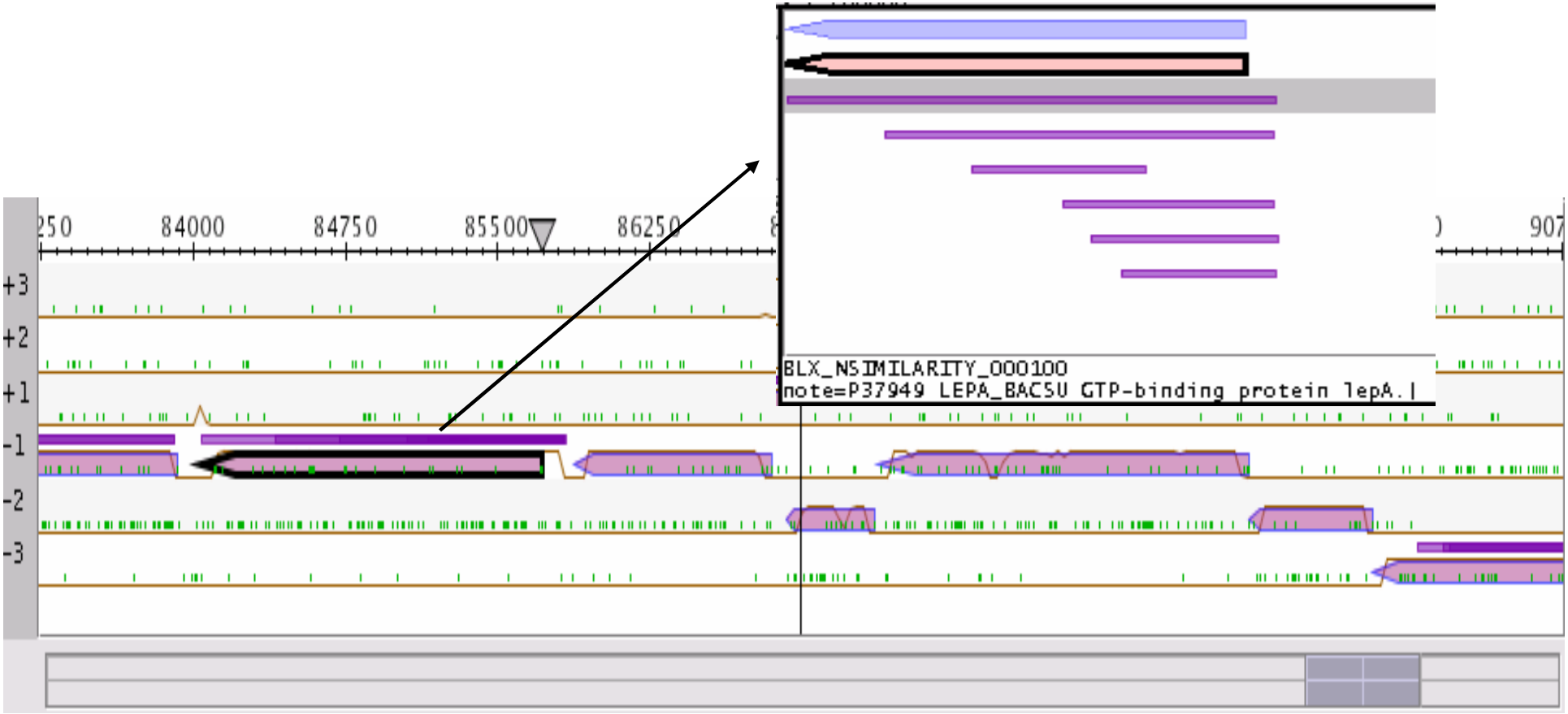
GenoAnnot : reannotation de génome (1)

Génome de *Borrelia b. :* fragment des premiers 100 kb [Fraser et al. 97]



GenoAnnot : reannotation de génome (2)

- Combinaison de résultats de plusieurs méthodes : i.e correction de start



Interoperabilité GenoAnnot -> GenoBool (1)

- Importation des CDS dans GenoBool
- Recherche de classes d'usage des codons (contingence + AFC)

File Rows Columns Graph Task Window

Import

Choose variat
GA_FeatCds
Instanc
gcPerce
gcClass
annotat
quantit
originIr
origin
qualitat
chemic
kingdo
history
Cancel

File Rows Columns Graph Task Window

Query CDS_gbk

	on_G...	startPositio...	seqFragment_GA_IsLoc...	endPosition_...	seqFragment_GA_IsLocatedOn12:AAA_K	seqFr
1		105	ATGATACTAATGAAGT...	677	13	1
2		105	ATGATACTAATGAAGT...	677	13	1
3		768	ATGGACTTTTTAAAAA...	1796	20	1
4		768	ATGGACTTTTTAAAAA...	1796	20	1
5		1784	TTGCGATACCATGAA...	3148	34	3
6		1784	TTGCGATACCATGAA...	3148	34	3
7		3395	ATGCTTAAACAATATT...	5188	60	19
8		3395	ATGCTTAAACAATATT...	5188	60	19
9		5251	TTGATTTTGAAAAGAA...	6312	25	2
10		5251	TTGATTTTGAAAAGAA...	6312	25	2
11		6309	TTGATAAAAGCAGTA...	7433	18	2
12		6309	TTGATAAAAGCAGTA...	7433	18	2
13		7458	ATGAATGAAAGAGAA...	8315	14	1
14		7458	ATGAATGAAAGAGAA...	8315	14	1
15		8412	TTGAGTAGAAAATTTA...	9197	20	3
16		8412	TTGAGTAGAAAATTTA...	9197	20	3
17		9202	ATGGAAATAGGCAAA...	10206	43	11
18		9202	ATGGAAATAGGCAAA...	10206	43	11
19		10203	ATGAAGTCAATAGGA...	10577	16	0
20		10203	ATGAAGTCAATAGGA...	10577	16	0
21		10581	ATGAAAAAAATATCA...	11420	39	10
22		10581	ATGAAAAAAATATCA...	11420	39	10

0 selected rows

Tasks

- BioCoders
  - Amino acid usage
  - Codon usage (1)
  - DiNucleotide usage
  - Nucleotide usage
- Analysis
  - ACM
  - ACM2
  - ACP
  - AFC (2)
- Clustering
  - Ether

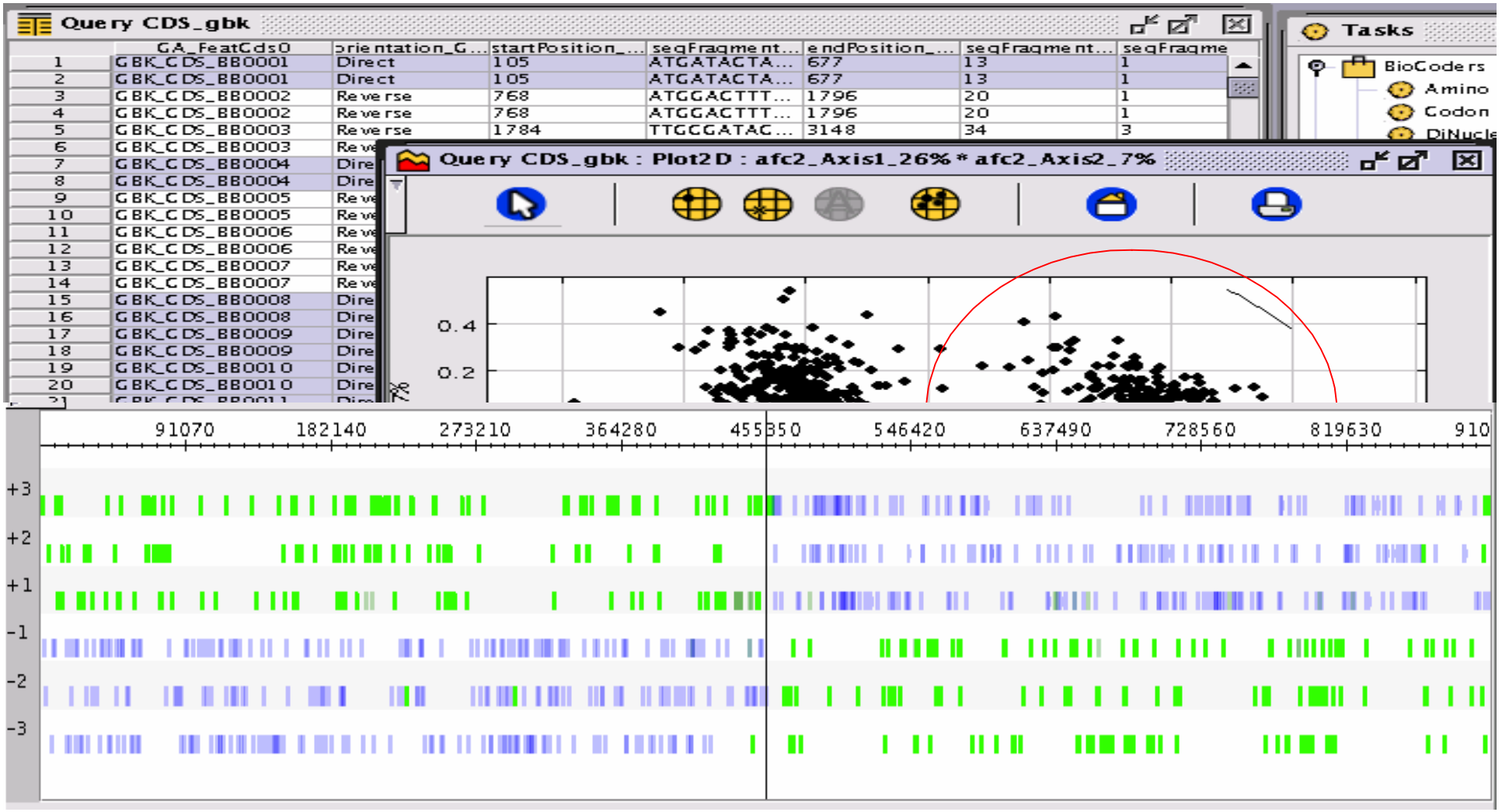
Codon usage

documentation:

Compute codon usage from a nucleotidic sequence

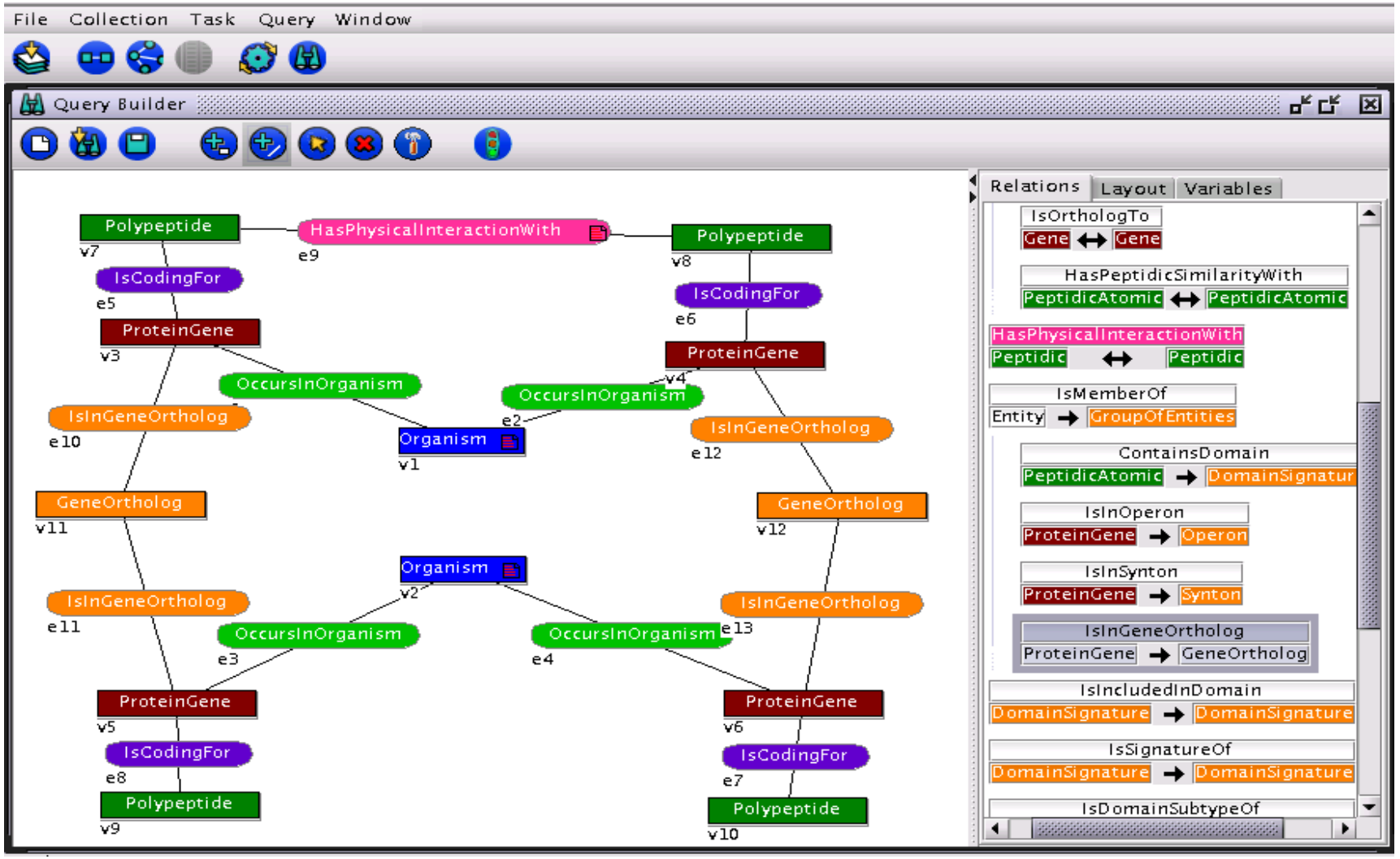
Interoperabilité GenoBool->GenoAnnot (2)

- Identification des sélectionnées CDS dans GenoAnnot
- Visualiser les CDS leading vs lagging

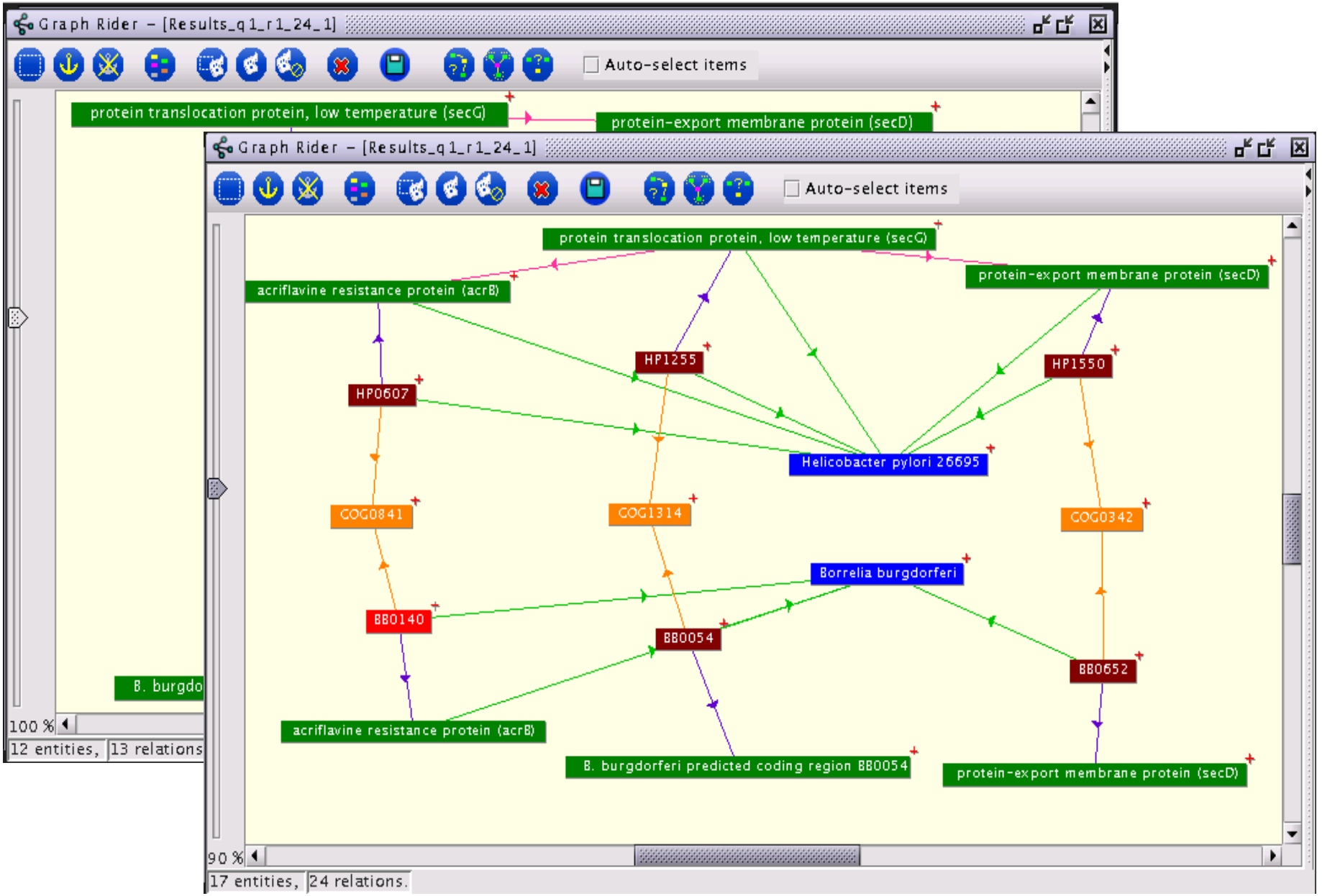


GenoLink : exploration du voisinage (1)

Espace de données : gènes, polypeptides, COGs, PIM pylori (Hgx)



GenoLink : exploration des voisinages (2)



- Genostar est mis à disposition des laboratoires de recherche publics

  Décembre 2002 : Version 1.1

  Juin 2003 : Version 1.2 (interopérabilité, ajout tâches…)

  Décembre 2003 : Version 2.0

  – programmable : ajout de méthodes, création de stratégie…

Actuellement en alpha-test puis beta-test

Pour en savoir plus…

http://www.genostar.org

Demande : academic@genostar.org , industry@genostar.org

Support / Retours utilisateurs : report@genostar.org

Le consortium Genostar

GENOME express

INSTITUT PASTEUR

Genostar

INRIA

hybrigenics

People & Acknowledgements

Christophe Bruley, INRIA, puis Genome express
Pierre-Emmanuel Ciron, INRIA
Antoine Danchin, Institut Pasteur
Stéphane Declere, Genome express
Jean-Louis Divol, Hybrigenics
Véronique Dupierris, INRIA
Patrick Durand, Hybrigenics
Gilles Faucherand, Genome express
Agnès Iltis, INRIA
Laurent Labarre, Hybrigenics
Claudine Medigue, Institut Pasteur
Alain Meil, Hybrigenics
Anne Morgat, INRIA
François Rechenmann, INRIA
Hélène Rivière-Rolland, Genome express
Vincent Schächter, Hybrigenics
Thierry Vermat, Genome express
Yves Vandenbrouck, Genome express
Alain Viari, INRIA
Jérome Wojick, Hybrigenics

Genostar