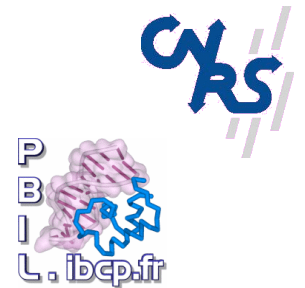


Une grille bioinformatique en France - expériences et perspectives «

Christophe Blanchet

PBIL IBCP , UMR5086 CNRS/UCBL
Pôle BioInformatique de Lyon
Institut de Biologie et Chimie des Protéines
Lyon-Gerland, France
Christophe.Blanchet@ibcp.fr

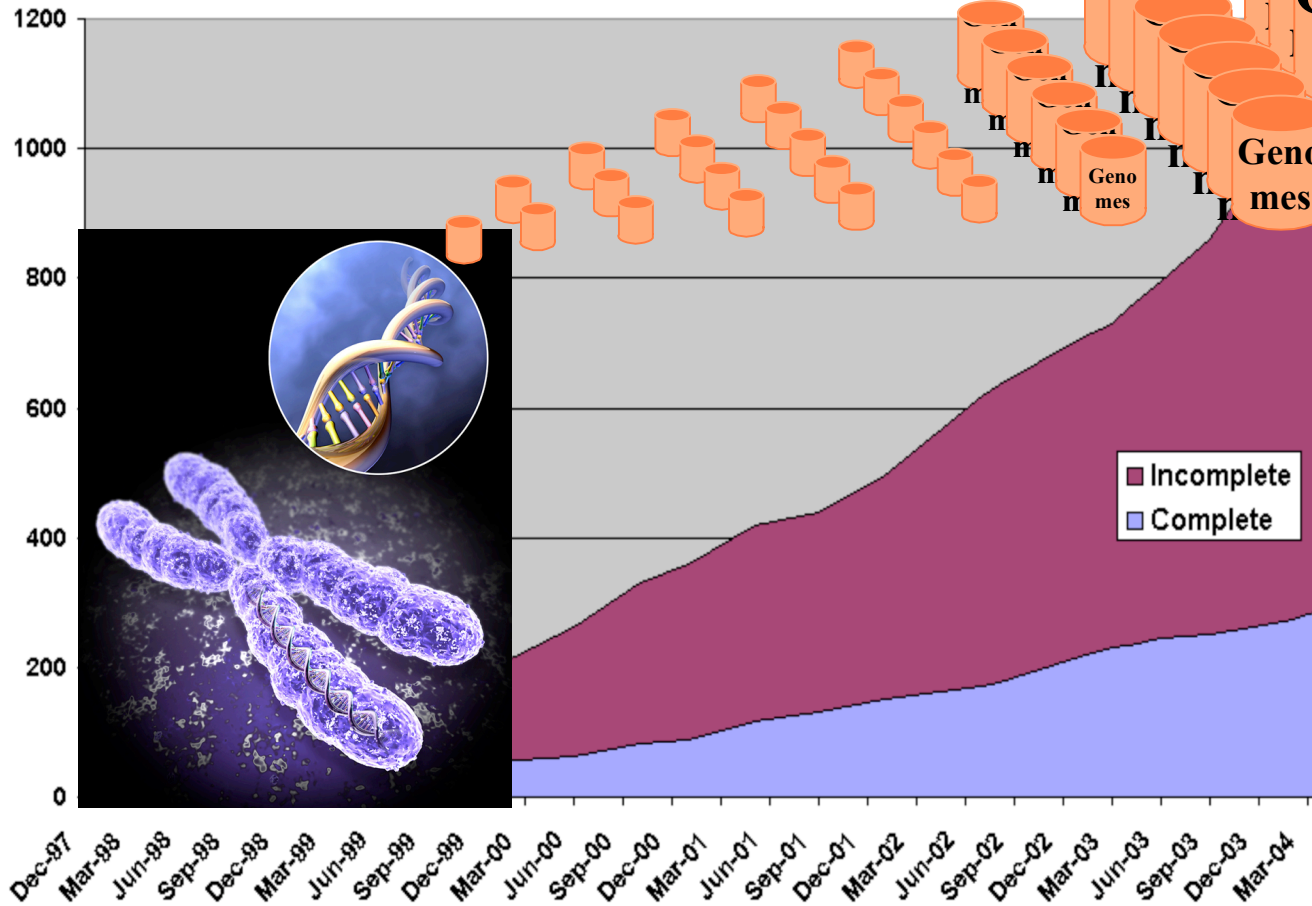


Contexte biologique

Bioinformatique à l'échelle des génomes

Complete genome era

Genome Sequencing Projects on GOLD ©



**Complete genome project: 1945
(Feb 2006)**

Complete (Published): 340

On going: 989 Prokaryotes, 588 Eukaryotes

e.g. of genome size

- *H sapiens*: 3.000 Mb (24)
- *S. cerevisiae*: 120 Mb,
- *A. thaliana*: 116 Mb

Ex. of Bioinformatics Applications

- **Different algorithms**

- Sequence similarity,
- Multiple alignment
- Structural prediction

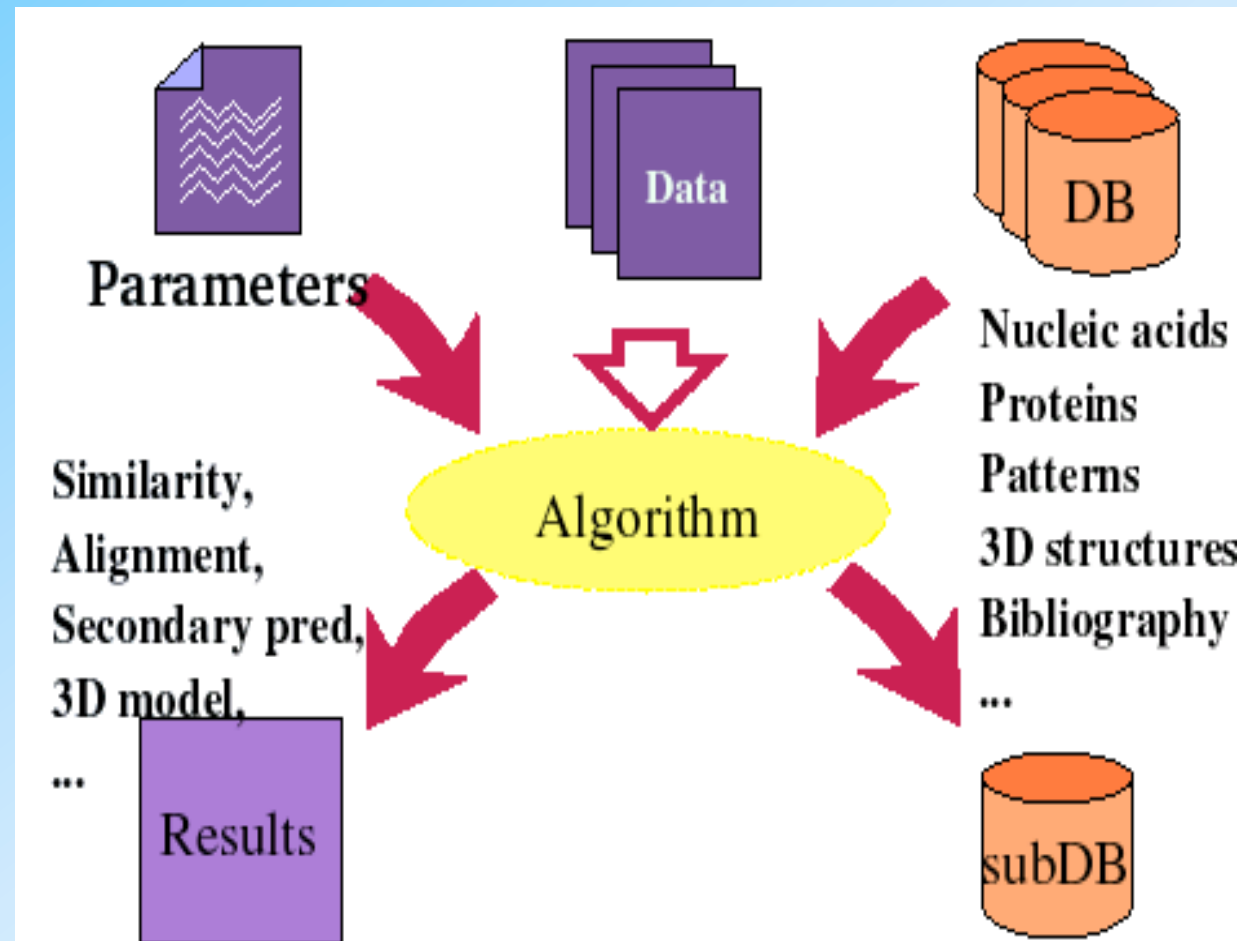
- **Numerous programs**

- BioCatalog:
 - + 600 at end of 1990s
- EMBOSS:
 - + 200 (world-famous)

- **Data access**

- Text files
- I/O standards with local file interface

No modification of source codes to preserve generic model



Besoins intensifs en Bionformatique

◦ **Calculs**

- Peuvent être satisfait s'ils sont peu liés aux données biologiques

◦ **Connaissances (données ou outils)**

- Besoins essentiels difficile à satisfaire de par la nature des données
 - *Taille intermédiaire*
 - *Complexité et hétérogénéité (sémantique)*
 - *Dispersion géographique des sources*



Expériences française en Bionformatique sur les grilles

Animation, conception, réalisations

GiGn

Grille pour la Génomique

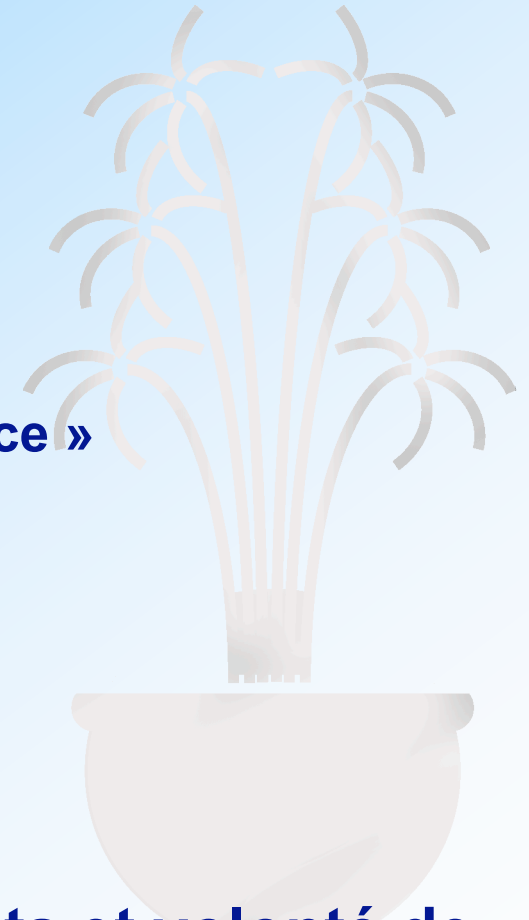
- **Groupe thématique national**
 - synergie entre les différents acteurs de la recherche en bioinformatique et en grille informatique au sein des communautés Biologie, Informatique, Mathématiques et Sciences Physiques
 - Soutenu par l'action IMPG
- **Animation de la communauté de 2003 à 2005**
 - 50 participants/réunion en moy.
- **<http://gign.ibcp.fr>**
- **Responsables**
 - C. Blanchet, V. Breton, D. Lavenier

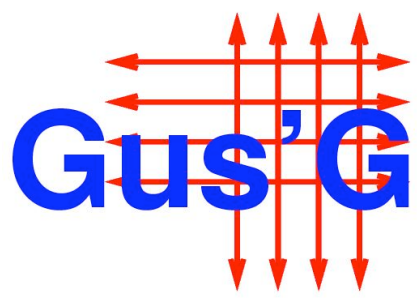


GiGn

Grille pour la Génomique

- **GiGn 2003**
 - 15 jan 2003, IBCP, Lyon
- **GiGn 2004**
 - 28 jan 2004, LPC, Clermont-Ferrand
- **OntoBIO - GiGn 2004**
 - En partenariat avec l'action spécifique OntoBIO
 - 20 oct 2004, Cines, Montpellier
- **Journée 2005 « Une grille Bioinformatique en France »**
 - En partenariat avec l'association X/Aristote
 - 16 juin 2005, IBCP, Lyon
- **OGSB 2005**
 - En partenariat avec l'ACI IsiBIO
 - 4 juil 2005, IBCP, Lyon
- **=> Intérêt notable de la part des participants et volonté de poursuivre la réflexion par des réalisations**



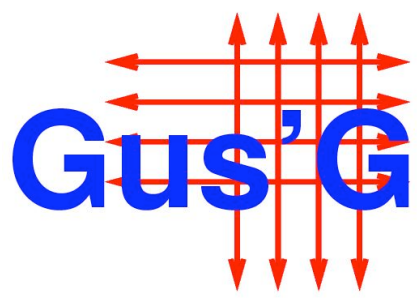


GusG : Grid User Group



- Groupe de travail de l'association X/Aristote
 - Sur l'usage des grilles par les communautés d'utilisateurs industriels et académiques
- Journée 2005 « Une grille Bioinformatique en France »
 - En partenariat avec GiGn
 - 16 juin 2005, IBCP, Lyon
- Journée 2005 sur les usages opérationnels et industriels des grilles
 - En partenariat avec EGEE Industry Forum
 - 17 juin 2005, IBCP, Lyon
- Responsables
 - P. d'Anfray, C. Blanchet



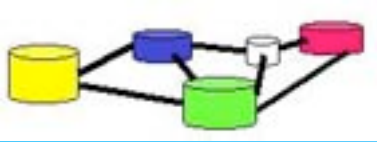


GusG : Grid User Group



- Annonce de la Journée 2006 de « Une grille Bioinformatique en France »
 - Poursuivre l'initiative lancée en 2003 avec GiGn
 - En partenariat avec **EMBRACE NoE** et **Décrypthon**
 - Prévue le 12 mai 2006, IBCP, Lyon
- Responsables
 - **C. Blanchet (IBCP), P. d'Anfray (Renater)**

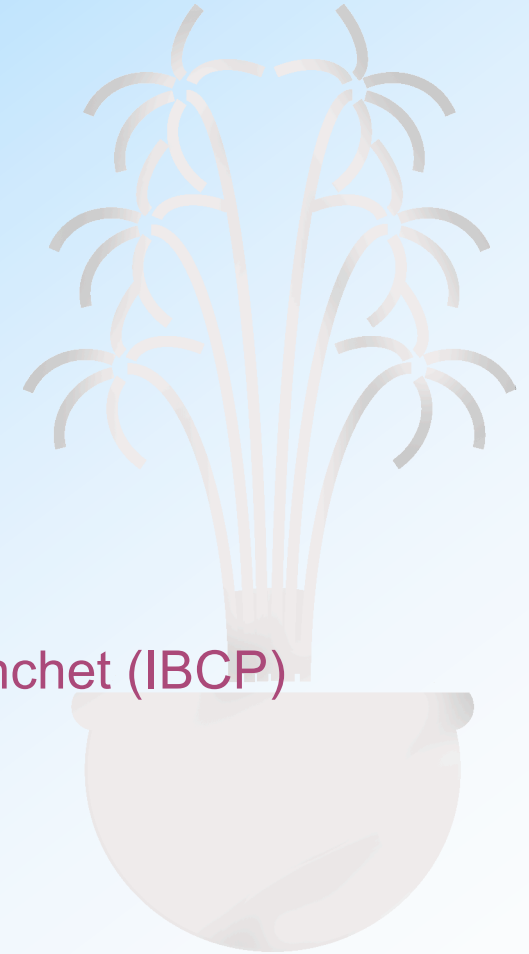




OGS-B

Ontology, Grid and Semantic integration for Biology

- **4 juillet 2005, IBCP, Lyon**
 - Réunion satellite du congrès JOBIM'05
- **Partenaires**
 - ACI IsiBIO
 - Action GiGnContacts:
 - [http:// gign.ibcp.fr/OGSB05](http://gign.ibcp.fr/OGSB05)
- **Responsables**
 - I. Mougnot (LIRMM), C. Froidevaux (LRI), C. Blanchet (IBCP)
- **=> Edition 2006**
 - Prévues le 4 juillet 2006
 - Réunion satellite de JOBIM 2006, Bordeaux



EMBRACE (EU FP6)

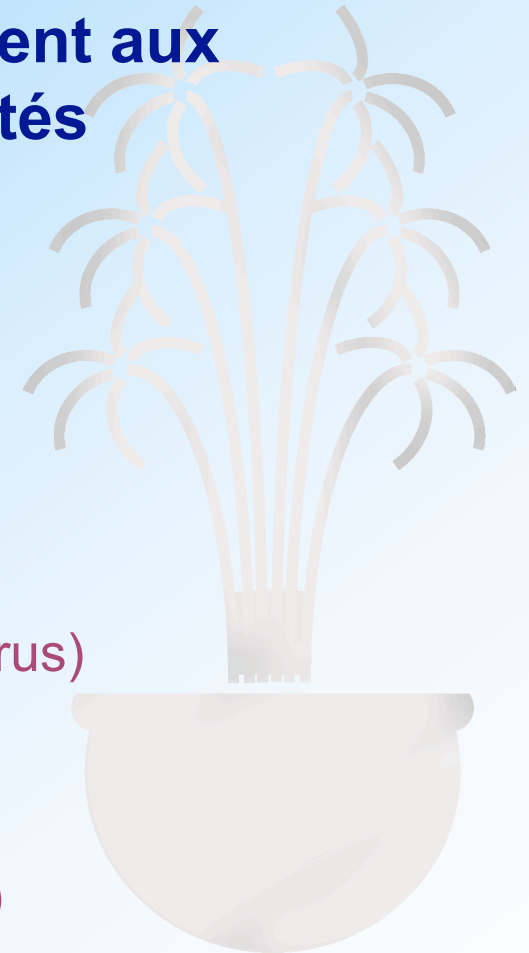


- « **A European Model for Bioinformatics Research and Community Education** »
- **Objectifs:**
 - Intégrer les données et les applications bioinformatiques dans un contexte de grille
 - Diffuser ces compétences à la communauté scientifique
- **Réseau d'Excellence Bioinformatique**
 - Depuis le 1er février 2005
 - Partenaires: EBI, EMBL, SIB, CNRS, INRA, Lion Bioscience, ...
 - PI pour le CNRS: C. Blanchet (IBCP)

EMBRACE: Participation CNRS



- Les partenaires français du CNRS participent aux comités scientifiques respectifs des activités
- **WP2.***
 - CNRS IBCP C. Blanchet (tool integration: NPS@)
- **WP3.***
 - CNRS LCP V. Breton (Responsable du WP)
- **WP4.***
 - CNRS IBCP C Combet (test cases : Hepatitis C Virus)
 - (INRA Toulouse) -> CNRS BBE, D Kahn
- **WP5.***
 - CNRS Tous (Outreach: conférences et formations)



EGEE Biomedical Activity

○ Chair:

- *Johan Montagnat , CNRS I3S*
- *Christophe Blanchet (deputy), CNRS IBCP*

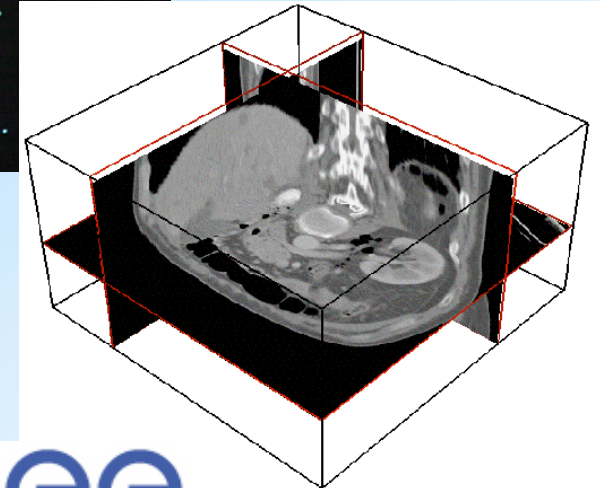
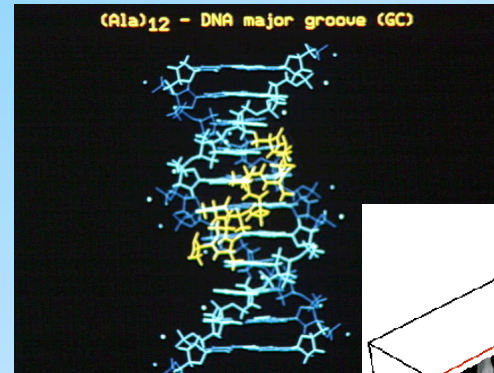
○ Biomedical activity area

- Bioinformatics
- Medical imaging
- Other health related areas

○ Three types of application

- **Pilots:** LCG-2 compliant applications at day 0
- **Internal:** from project partners, to be deployed on EGEE
- **External:** from other projects, to go through a selection procedure

- *EGEE User Forum, CERN, March 1-3th, 2006*
<http://egee-intranet.web.cern.ch/egee-intranet/User-Forum>



EGEE
Enabling Grids
for E-science

Une grille bioinformatique en France

Perspectives

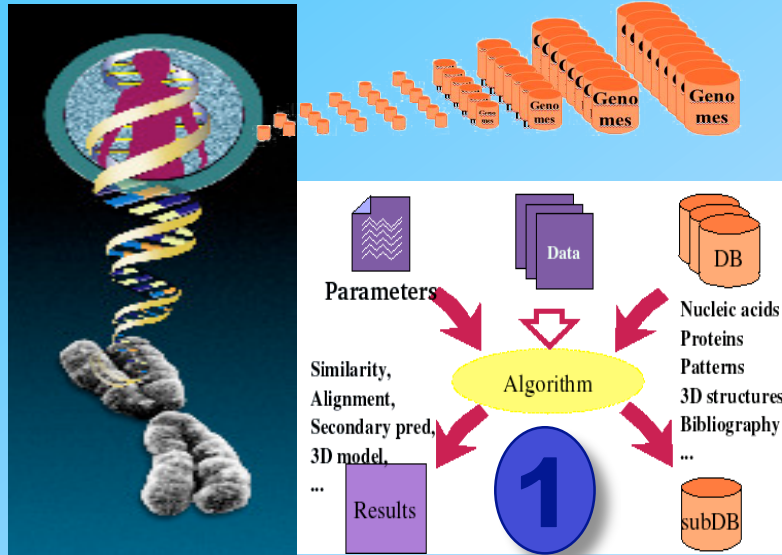
Objectifs

- **Fédérer la communauté bioinformatique française autour de ressources mutualisées**
 - Les données biologiques: intégrer les banques de données et les diffuser.
 - Le calcul en bioinformatique: intégrer les algorithmes et les rendre accessible.
- **Une organisation virtuelle (VO)**
 - Une identité: « GrilBio »
 - Mutualiser les ressources matérielles et scientifiques
 - Définir et appliquer une politique d'accès et d'utilisation de ces ressources

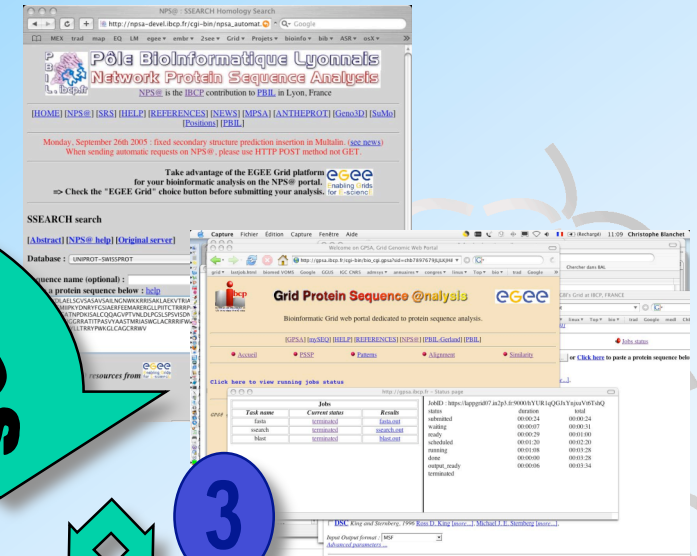
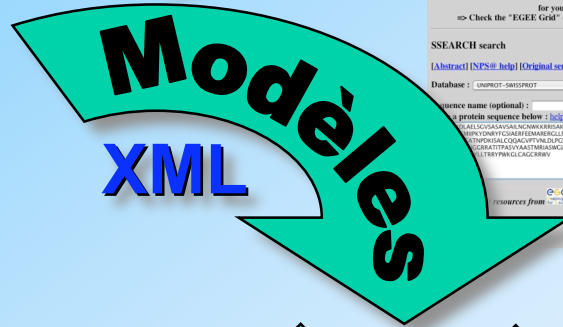


Distribuer les connaissances biologiques avec la grille

- Intégrer les données et les outils -



2



3

1) Modéliser les données biologiques et les outils bioinformatiques

- Modèle des banques: Swiss-Prot, TrEMBL, PROSITE, ...
- Modèle des données/métadonnées : protéines, gènes, ...
- Modèles des outils: I/O, paramètres, exécution, distribution, utilisations des données, ...

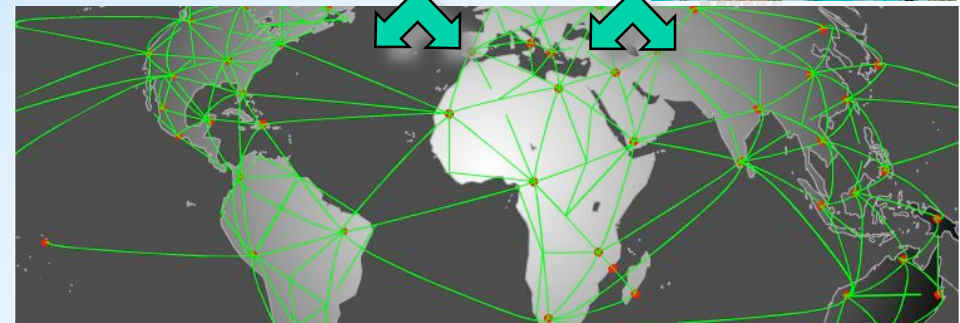
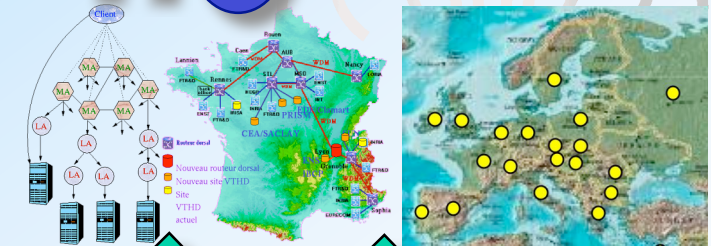
2) Formaliser les modèles avec le langage XML. Connecter ces modèles en processus complexes.

- DTD « bio_methods »
- Descripteurs XML d'outils: blast.xml, ssearch.xml, clustalw.xml, pattinprot.xml, sopm.xml, phd.xml, ...
- Descripteurs XML de plate-formes: EGEE, RUGBI, ...

3) Intégrer ces données et outils dans des plateformes informatiques distribuées

- Banques de données: Swiss-Prot, TrEMBL, PROSITE
- Méthodes: BLAST, FastA, Ssearch, ClustalW, Multalin, PatInProt, GOR4, PHD, SOPM, ...
- Plateformes distribuées: EGEE (EU), e-Toile (Fr), Grid5000 (Fr), ...

GRIDS:
EGEE (EU)
E-Toile (Fr)
DIET (Fr)



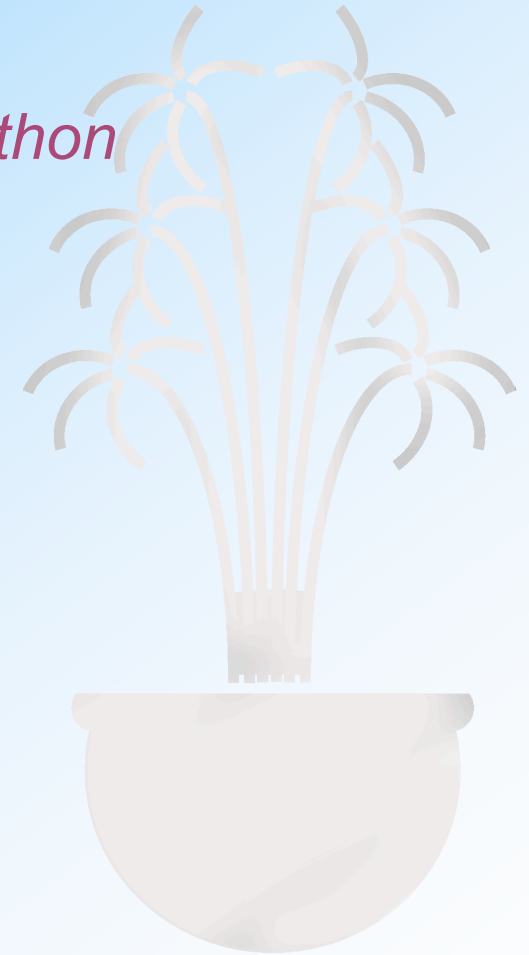
Plateformes de grille en France

◦ 3 types:

- Recherche en informatique: GdX, Grid5000
- Application à vocation académique:
 - *Opérationnelle: EGEE, DEISA, Décryphon*
 - *Expérimentale: GénoGrid*
- Application à vocation industrielle
 - *RUGBI, fr-grid*

◦ Participation des CCs ?

- EGEE: CC-IN2P3, CINES
- DEISA: IDRIS
- ?? : CEA



Une Grille Bioinformatique en France ?

- **Grille dédiée ou pluridisciplinaire ?**
 - « Dédiée » serait inefficace et coûteux.
 - Définir la participation de la communauté bioinformatique à une grille pluri-disciplinaire
- **Quelle plate-forme (intergiciel) ?**
 - Grid5000
 - DEISA
 - EGEE
- **Quels moyens humains pour la VO « GrilBio » ?**
 - Coordination
 - Animation de la communauté
 - Noyau d'experts
 - *Conseil et assistance à la communauté*
 - *Interlocuteur avec les « plate-formes » de grille*



Une Grille Bioinformatique en France ?

- **Volonté des organismes de tutelle ? CNRS ?**
- **Collaboration des nombreux pôles d'excellence en Bioinformatique:**
 - Strasbourg, Paris, Lyon, Toulouse, Rennes, Marseille, ...
 - Compétences complémentaires
- **Modèle monolithique ne semble pas avoir fonctionné ...**
- **A concerter avec les initiatives nationales:**
 - Génopoles, ReNaBi, Décrypthon,

