

# PROJET AGENAE

Analyse des **GEN**omes des **Animaux d'Elevage**

- Le programme **AGENAE** :
  - \* développer des recherches génériques et des actions de recherche finalisées dans le domaine de la génomique animale.
  - \* quatre espèces majeures (bovin, porc, poule et truite),
  - \* cartographie des génomes
  - \* expression génétique.
- Un **Groupement d'Intérêt Scientifique** a été mis en place en 2002, dirigé par Claude Chevalet (INRA Toulouse).
  - \* regroupe l'INRA, le CIRAD
  - \* les organismes interprofessionnels et partenaires des filières

# PROJET AGENAE

Analyse des **GEN**omes des **Animaux** d'Élevage

## x 7 axes

- **Caractérisation des transcriptomes**
- **Etudes d'expression des gènes**
- **Cartographie et Structure des Génomes**
- **Programmes finalisés en Génétique**
- **Bioinformatique** : un groupe bioinformatique (SIGENAE) et des recherches en cartographie comparée
- **Infrastructures** : un centre national de ressources (GABIE) et des plates-formes utilisatrices à l'INRA (Toulouse, Jouy, Rennes, Tours, Clermont)
- **Animation - Communication** : un groupe veille scientifique, un groupe communication

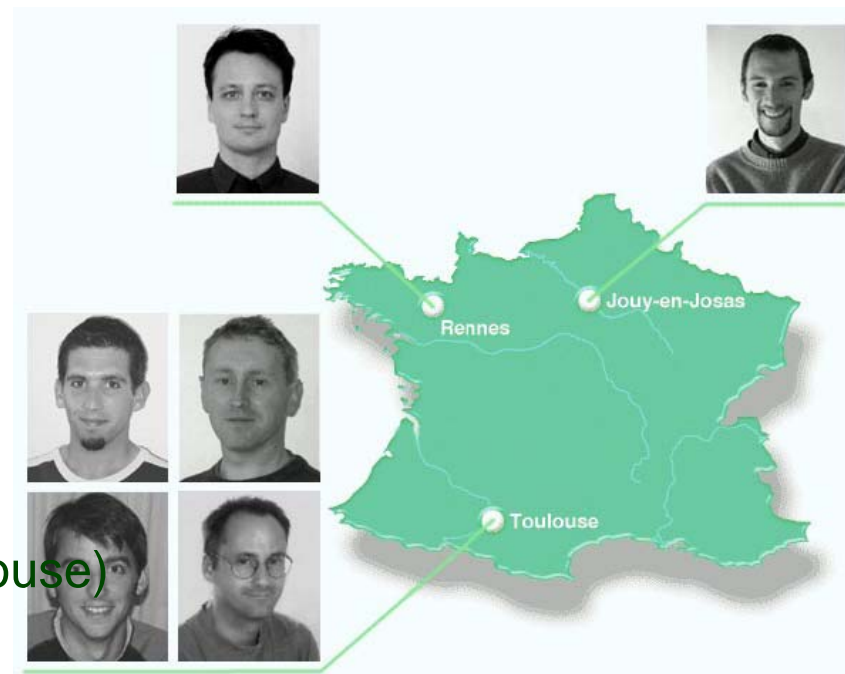
# SIGENAE

## Système d'Information du projet d'Analyse des **GEN**omes des **Animaux d'Elevage**

### × L'équipe

#### × 6 ingénieurs : 5,6 ETP

- Philippe Bardou (Toulouse)
- Patrice Dehais (Toulouse)
- Romain Freme (Toulouse CDD)
- François Moreews (Rennes)
- Cédric Cabau (Jouy-en-Josas)
- Animation : Christophe Klopp(Toulouse)



# Les missions dans AGENAE

Au départ :

- × Mise en place de bases de données pour les génomes des espèces animales
- × Acquisition et intégration continue de compétences
- × Renforcement de l'effort de recherche méthodologique

Maintenant :

- × Service aux utilisateurs sur 2 types de données :
  - séquences
  - puces à ADN

# Autres projets

- × projet européen EADGENE
  - Animation de WP : bio-informatic facilities
- × prestation de service :
  - Aquafirst : (séquences / bio-puces)
  - Aquagenome et Aquafunc : expertise bio-informatique
- × partenariat avec le RNG
  - Toulouse-Genopole
  - Ouest-Genopole

# Particularités

- ★ inter-département (Phase, GA, SA, MIA)
  - ★ hébergement externe
  - ★ équipe de service
  - ★ pas une unité au sens INRA
  - ★ pas de machine propre
  - ★ stratégie dirigée vers l'extérieur
- missions internes et externes à l'INRA

# Les machines

## Accès web



Serveur Web (4 cpu / 8Go / 1To)



Cluster de calcul (10 noeuds / bipro 4Go)



Serveur de Base de données (4 cpu / 8 Go )



Serveur de développement (4 cpu / 8 Go)

# L'offre de service

## × Séquences :

- nettoyage, assemblage, annotation, publication
- interrogation, récupération

## × Puces à ADN

- traitement d'image
- stockage des données de quantification
- traitement statistique des données et interprétation des résultats

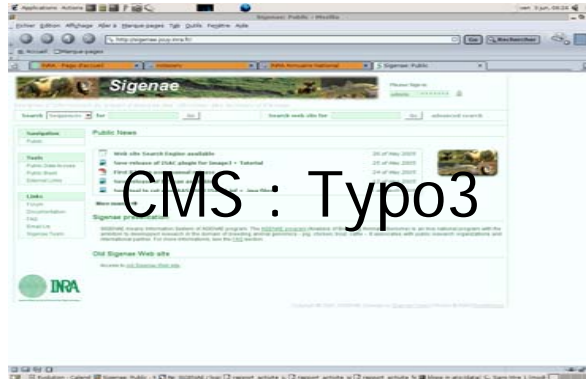
## × hébergement de site web



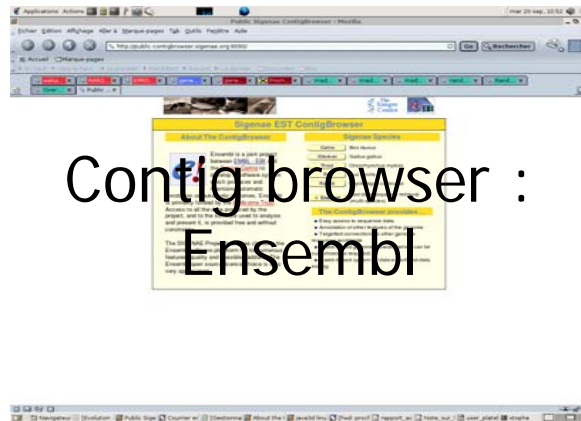
# Stratégie

- × Utilisation et modification de logiciels existants
  - conformité au besoin
  - communauté d'utilisateurs et de développeurs
- × Utilisation de logiciels libres
  - pérennité
  - accès aux sources en cas de modification
  - participation à la communauté

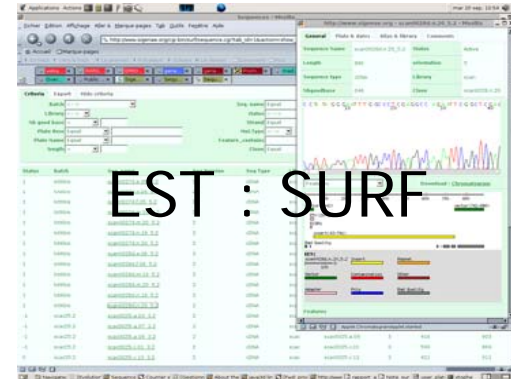
# Applications Serveur



CMS : Typo3



Contig browser :  
Ensembl

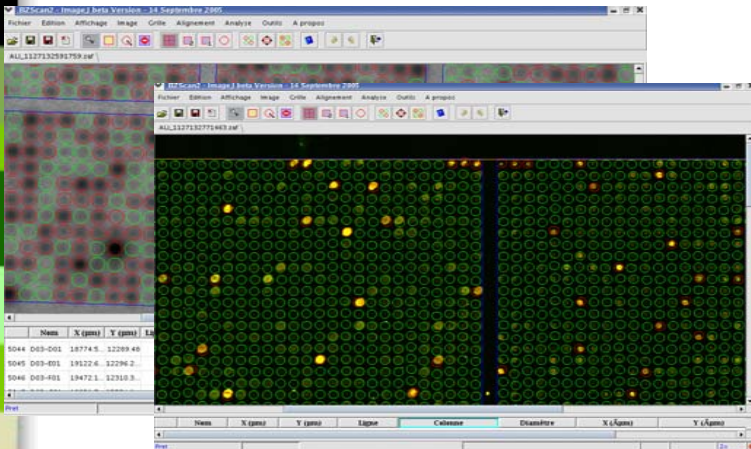
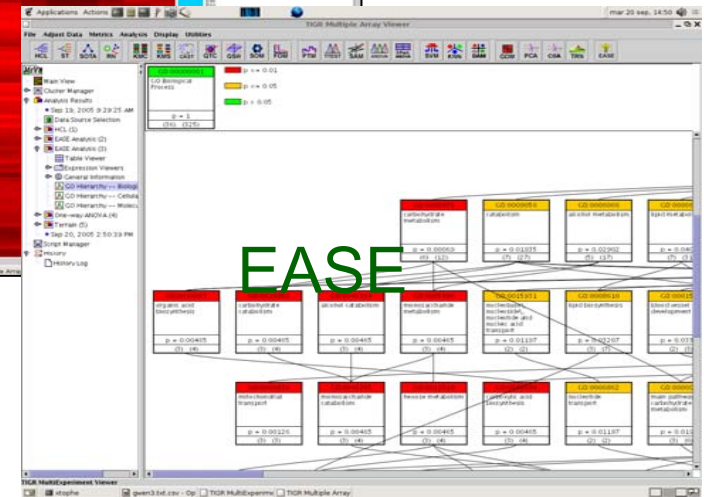
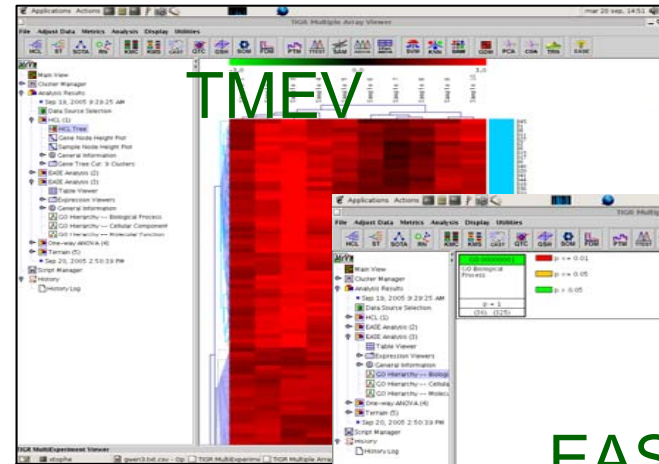
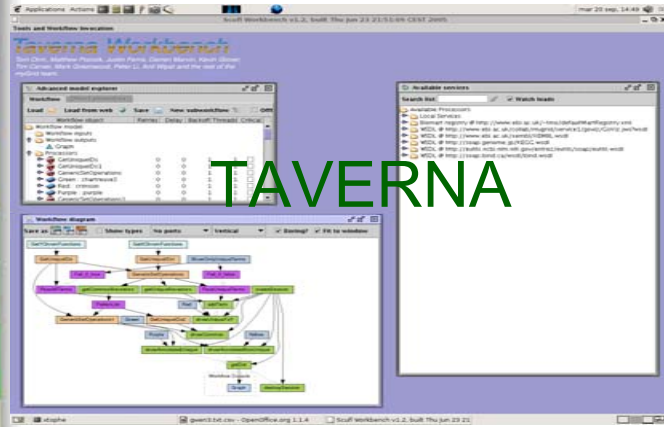


EST : SURF

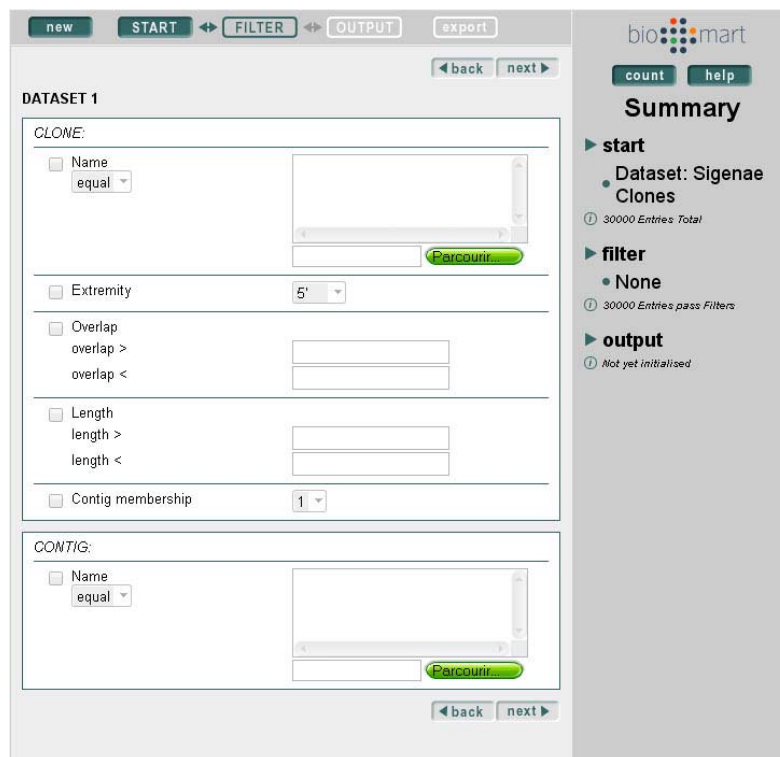


Microarray :  
BASE

# Applications Client



# Ensembl / BioMart



The screenshot shows the BioMart interface. At the top, there are navigation buttons: 'new', 'START', 'FILTER', 'OUTPUT', and 'export'. Below this, there are 'back' and 'next' buttons. The main area is divided into two sections: 'DATASET 1' and 'Summary'.

**DATASET 1**

**CLONE:**

- Name: equal
- Extremity: 5'
- Overlap: overlap >, overlap <
- Length: length >, length <
- Contig membership: 1

**CONTIG:**

- Name: equal

**Summary**

- start**
  - Dataset: Sigenae Clones
  - 30000 Entries Total
- filter**
  - None
  - 30000 Entries pass Filters
- output**
  - Not yet initialized

Biomart : generateur d'interfaces de requete a une bd SQL




Ensembl contig browser

# BASE

## BioArray Software Environment

### x gestion des données de microarrays et publication



Logged in as root  
[Log out]  
Superuser  
Users online: 1 [View]

- Reporters
- Array LIMS**
- Plate types
- Plates
- Reporter search
- Array designs
- Array batches
- Array slides
- Plate file formats
- Plate mapping formats
- Reporter map formats
- Protocols

**array slides**

Field	Op	Value	Buttons	Translated value
-	=		Add/Update	

Presets Save current as new preset

<<prev next>> 1 2 3 4 5 6 10 20 30 36 (3596 hits, 100 per page)

Barcode	Array batch	Ordinal	Array design
TroutGeneric2 - 139 - 5-2	TroutGeneric2 - 139 - 5 2		TroutGeneric
TroutGeneric2 - 139 - 5-3	TroutGeneric2 - 139 - 5 3		TroutGeneric
TroutGeneric2 - 139 - 5-4	TroutGeneric2 - 139 - 5 4		TroutGeneric
TroutGeneric2 - 139 - 5-5	TroutGeneric2 - 139 - 5 5		TroutGeneric
TroutGeneric2 - 139 - 5-6	TroutGeneric2 - 139 - 5 6		TroutGeneric
TroutGeneric2 - 139 - 5-7	TroutGeneric2 - 139 - 5 7		TroutGeneric
TroutGeneric2 - 139 - 5-8	TroutGeneric2 - 139 - 5 8		TroutGeneric
TroutGeneric2 - 139 - 5-9	TroutGeneric2 - 139 - 5 9		TroutGeneric
TroutGeneric2 - 139 - 5-10	TroutGeneric2 - 139 - 5 10		TroutGeneric
TroutGeneric2 - 139 - 5-11	TroutGeneric2 - 139 - 5 11		TroutGeneric

**Select the items you want to export in SOFT format**

Submission type:

Experiment (SOFT series)  Raw data sets (SOFT samples)  Array design (SOFT platform)

Raw data set(s):

Trout\_ForGEO\_2 (owner : root )

Trout\_ForGEO (owner : root )

Parameters:

GEO user ID

GEO password

Platform distribution

Platform manufacturer

Platform manufacture protocol  (if no configuration protocole)

Institute


Web link

Sample clone ID description

Normalized values extracted from:  BASE BioAssaySet (name like: )

New submission  or update

## GEO exporter

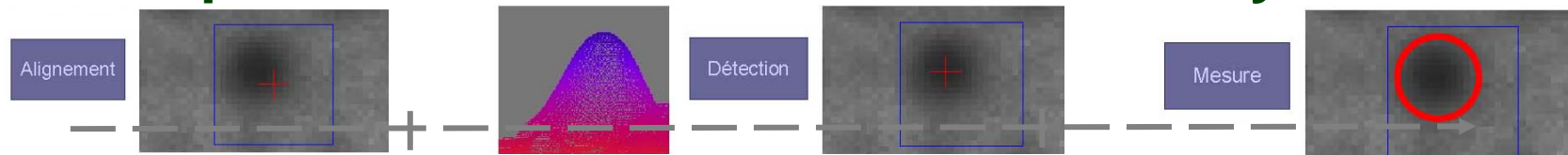


SIGENAE

octobre 2006

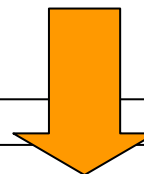
13

# AGScan : un logiciel libre, évolutif et automatique de quantification d'images de microarrays



## Adaptabilité nécessaire pour gérer :

- Diversité des puces : spots, canaux, marqueurs, hybridation
- Evolution technique constante



## Besoin d'une architecture logiciel permettant une évolutivité maximale :

- N canaux (pile)
- Grilles configurables
- Algorithmes
- langues

Architecture à  
base de **plug-ins**

# Taverna

diffusion de  
scenarii  
d'annotation

output	input	file	description
gene locations on caryotype view	gene Id, species, chromosome	<a href="#">KaryoviewFromGenesIdentifiers.xml</a>	display genes on a caryotype from gene identifiers, using Ensembl Karyoview
fasta sequence	DDBJ/EMBL/GenBank mRNA ID	<a href="#">SIGENAE_extract_cds_fasta.xml</a>	extract coding sequences (CDS) from remote EBI databases
alignment	sequence	<a href="#">BlastAtEBIFromSeq.xml</a>	NCBI Blast2 and WU BLASTat EBI on nucleic and proteic data banks
Gene Ontology term diagram	key word or GO ID	<a href="#">GOcontext.xml</a>	Gene Ontology
alignment	nucleic sequence	<a href="#">NarcisseQuery.xml</a>	BLAST hits sorted by species
		<a href="#">(with the possibility of Blast2NuclFromCloneName.xml)</a>	complex workflow clone -> Sigenae contig -> sequence -> alignment (name)
	record by line)	<a href="#">InputListGenerator.xml</a>	Taverna input list generator
	quence EMBL ID	<a href="#">BlastAtEBIFromEMBLID.xml</a>	NCBI Blast 2 and WU Blast at EBI from EMBL sequence ID
	quence	<a href="#">Blast2NCBlastNCBInucleicFromSeq.xml</a>	Blast2 at NCBI
	ID(NCBI) or URL(EBI)	<a href="#">retrieveBlastResultFromResultID.xml</a>	get Blast result from Blast Result ID(NCBI) or URL(EBI)
	quence EMBL ID	<a href="#">getFastaSequenceFromEMBLID.xml</a>	get FASTA sequence from ID



traitement par lots

prototypage

Web Services

diffusion, échange  
communautaire

BIOMART

annotation à  
façon