



Les bases de données en génomique

Christophe Klopp

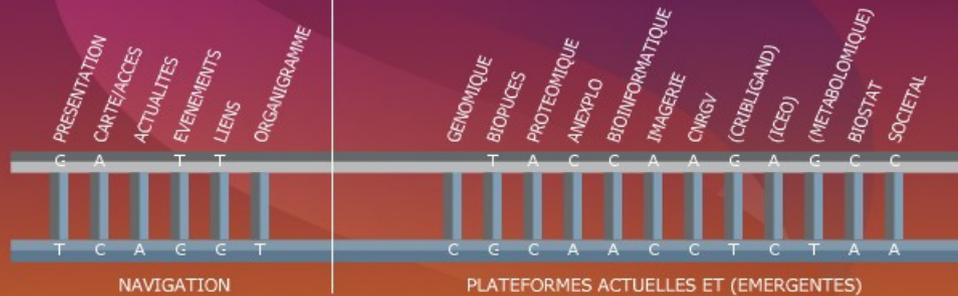
responsable technique de la plate-forme bio-info Genotoul

Le GIS Genotoul

Un réseau de plateformes
en sciences du vivant



GénoToul Toulouse Midi-Pyrénées
www.genotoul.fr



Bienvenue

Présentation
Nous contacter
Problème d'affichage ?

Actualités

20 oct 09 - Journée bioinfo : 3 Decembre !
28 sep 09 - GenoToul 2009 : 17 novembre !
15 mai 09 - Séminaire MIBS - 18 JUIN 2009

Plus ...



La plate-forme bio-informatique

short english presentation 



- La plateforme
- Banques
- Logiciels
- Infrastructure
- Formations
- Aides
- Retour à l'accueil



Plateforme bioinformatique

GenoToul

Chercher

Recherche avancée

Bioinfo-Genotoul Accueil

Bienvenue sur la plate-forme bio-informatique genotoul !

La plate-forme bio-informatique est l'une des douze plate-formes du **GIS Genotoul**. Elle a été labellisée RIO respectivement en 2003 et 2006. Depuis 2007, elle bénéficie du label IBISA. Son domaine d'activité est centré sur l'analyse de séquences.

Les prestations offertes incluent l'accès à :

- la puissance de calcul apportée par les **serveurs** et le **cluster** de 350 coeurs acquis en 2009 ; l'accès à cette infrastructure nécessite de disposer d'un compte utilisateur (**demandez l'ouverture d'un compte utilisateur**) ;
- des **banques de données généralistes** ;
- des **logiciels bio-informatiques de base et spécifiques** ;
- l'hébergement de projets (demandez l'hébergement d'un projet) ;
- la **formation** ;
- l'**accompagnement** de vos projets de traitement de données ou de développement d'outils.

L'expertise disponible et l'analyse des besoins actuels nous ont conduit à investir de manière prioritaire dans le traitement des données issues des nouvelles techniques de séquençage et sur la détection et l'annotation des ARN non codant (**RNASpace.org**).

Depuis 2008, la plate-forme est engagée dans un partenariat avec la **plate-forme génomique** pour le traitement et la mise à disposition des séquences produites localement.

Accès aux ressources



Les données en bio-informatique

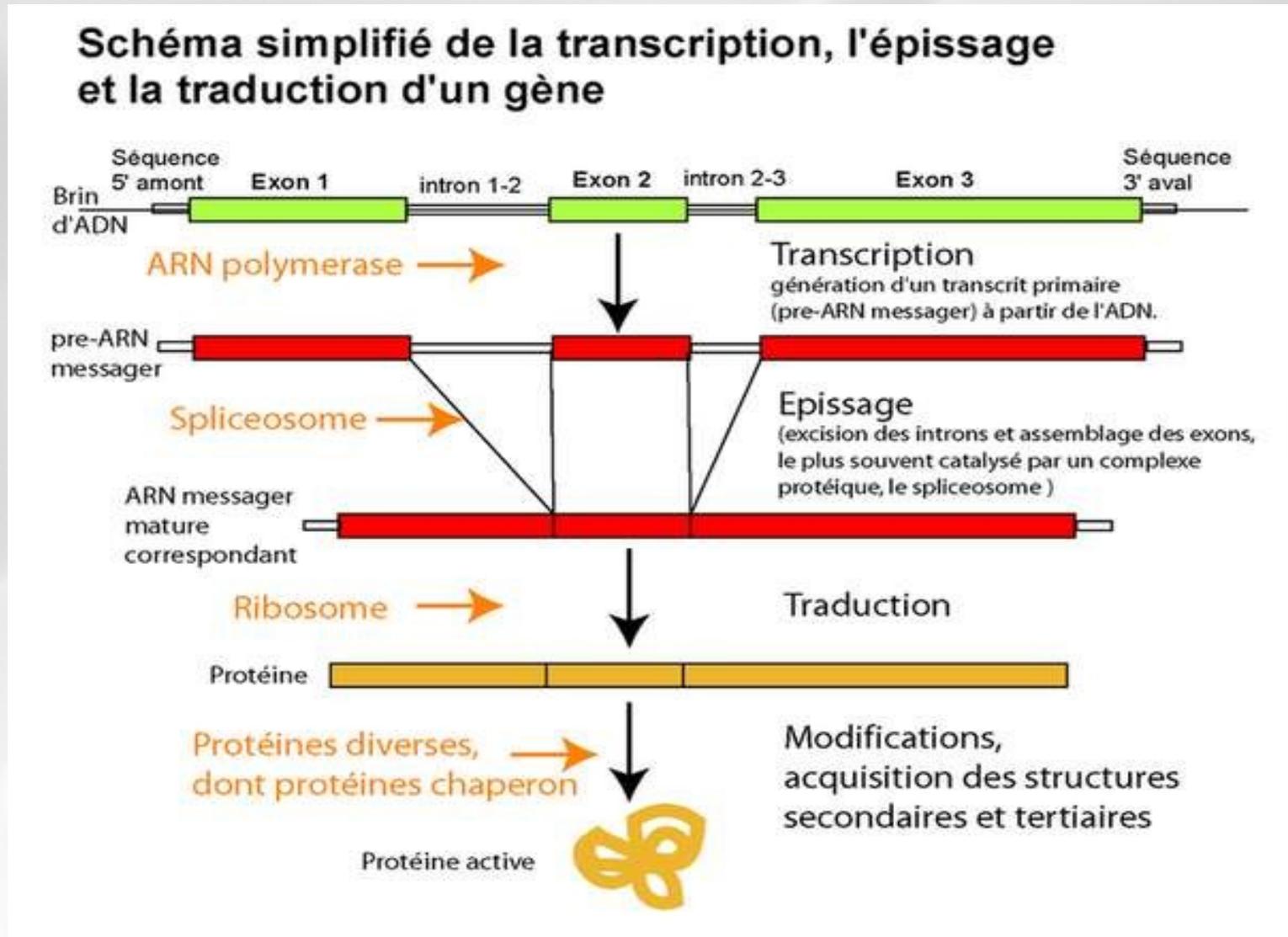
Wikipedia : La Bio-informatique est un champ de recherche multi-disciplinaire où travaillent de concert biologistes, informaticiens, mathématiciens et physiciens, dans le but de résoudre un problème scientifique posé par la biologie.

Données :

- Séquences nucléiques ou protéiques
- Polymorphisme et génotypage
- Intensité de signal (puce à ADN)
- Images (phénotype,...)
- ...

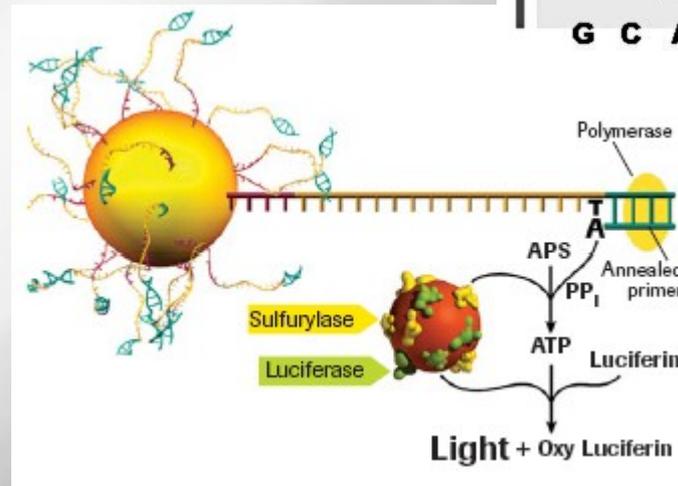
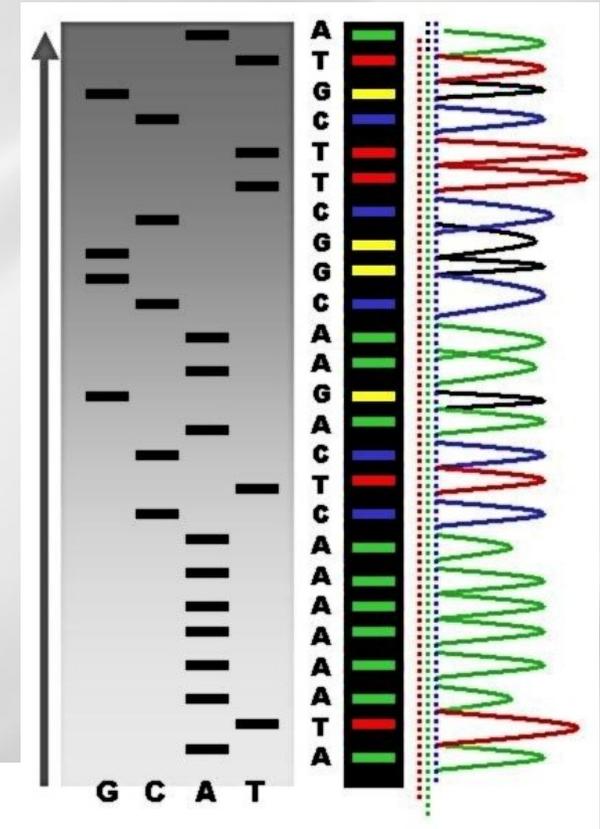
Les séquences

- Dogme de la biologie moléculaire :



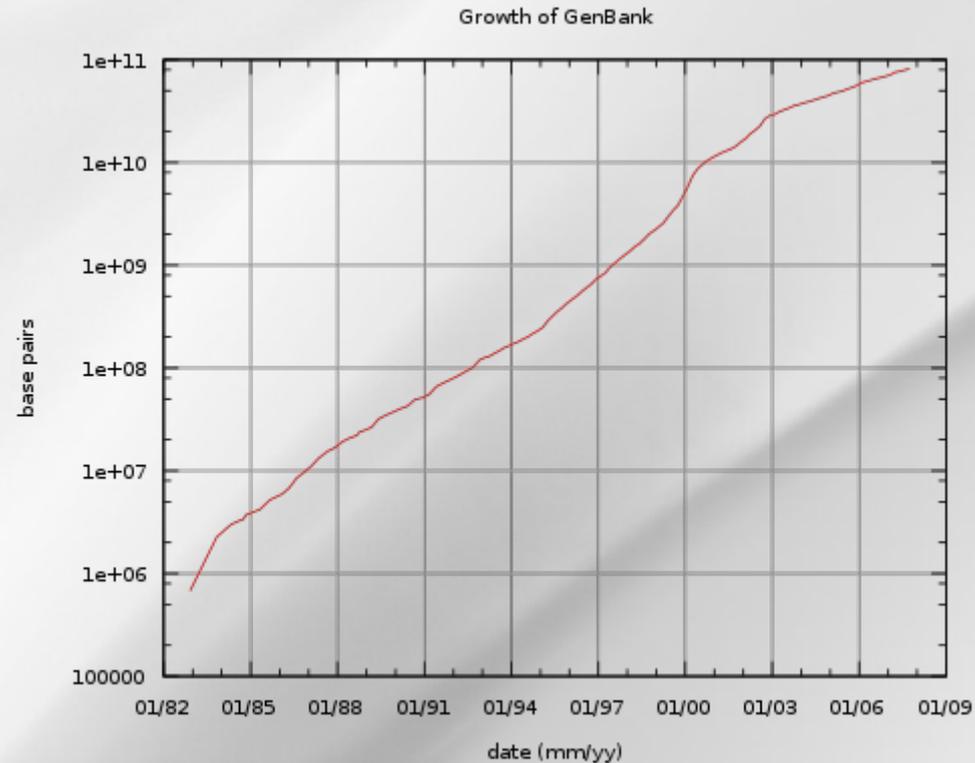
La production de séquences

- Méthode Sanger : 1970
- Nouvelles technologies :
 - Roche 454
 - Illumina Solexa
 - ABI Solid



Le partage des séquences

- Données publiques (Genbank)
- Partage entre différents centres : NCBI, EMBL, DDBJ



Les méta-données

- Informations sur
 - L'organisme
 - Le condition de production
 - Le producteur
 - ...

IDENTIFIERS

dbEST Id: 44992745
 EST name: CR376592
 GenBank Acc: CR376592
 GenBank gi: 126026477

CLONE INFO

Clone Id: tcbi0014.j.21 (5')
 Plate: tcbi0014c Row: j Column: 21
 DNA type: cDNA

PRIMERS

Sequencing: T7
 PolyA Tail: Unknown

SEQUENCE

```
ACGAGT CCCCCAT GAT GACCAGAGCT AT CCAT CCAT CT AGCCAGCCT CAACCAGAGT CC
GACACCCACT GAGGGCCT GCAT GGGGCT AGCT ACT GCT CAGAT GAAT GATTT CAAAT GGG
TTTT GGATT AGATTTTT CAAT GTTTTT AT GCT GT ACT GAT GCAACCCT GGT CT CAAAGG
CT CCGAGAAGCAGCAAGACT CCT GCACCT CCAT CAAT CAT GT CAT GTT CCACT GGT TTA
TT AAT ATT CT CCCAGT GT AAAT AAT GGAAGAT GCAT ATTTT GAAAGT GCACTTT CT TTTT
AAGT GGT ACAGT ATTT GCCAAAT CT CCGCCATT GCT AT GAT GT GAGAT CACAT CT CT GT
TAGACAGACAGACAT ACCAAAT GAACT GAGGGAATT AGAATT GGAACAGAAGT AGAAT AG
TT GGT GT ATTTTT ACAAT GTTTTT CT TT ACCAGCT GTTTT CACCT GCCAATTTT CACTT
AAAGT GCCTT GACAAGAACAT GGAAT GTTTTT AT AAGT CAGT GT GAT ACCGGT GAT ATT AG
TAATTCCACATTTT GTT ACAGCCT GAATT AAAAAAT GGATT AAAT AT AAACAAAAT CNNN
NNAN
```

Entry Created: Apr 22 2004
 Last Updated: Feb 21 2007

COMMENTS

Sequence cleaned of vector, adaptor and repetitions.
 Contact us at sigenasupport@jouy.inra.fr to obtain the chromatogram of this sequence.

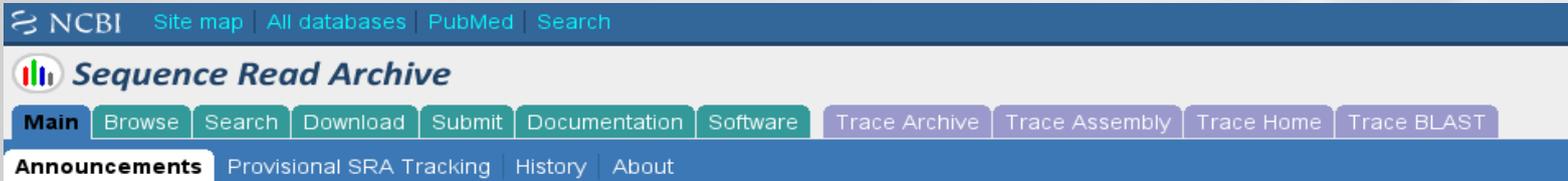
LIBRARY

Lib Name: AGENAE Rainbow trout normalized testis library (tcbi)
 Organism: [Oncorhynchus mykiss](#)
 Tissue type: testis
 Lab host: DH10B
 Vector: pT7T3D-PacI
 Description: Clone distribution : AGENAE Resource centre. Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73

SUBMITTER

Name: Guiguen Y
 Institution: INRA - SCRIBE
 Address: Campus de beaulieu, 35042 RENNES cedex, France
 Tel: 02.23.48.50.09
 Fax: 02.23.48.50.20
 E-mail: yann.guiguen@rennes.inra.fr

➤ Méta données par échantillon



NCBI Site map | All databases | PubMed | Search

Sequence Read Archive

Main | Browse | Search | Download | Submit | Documentation | Software | Trace Archive | Trace Assembly | Trace Home | Trace BLAST

Announcements | Provisional SRA Tracking | History | About

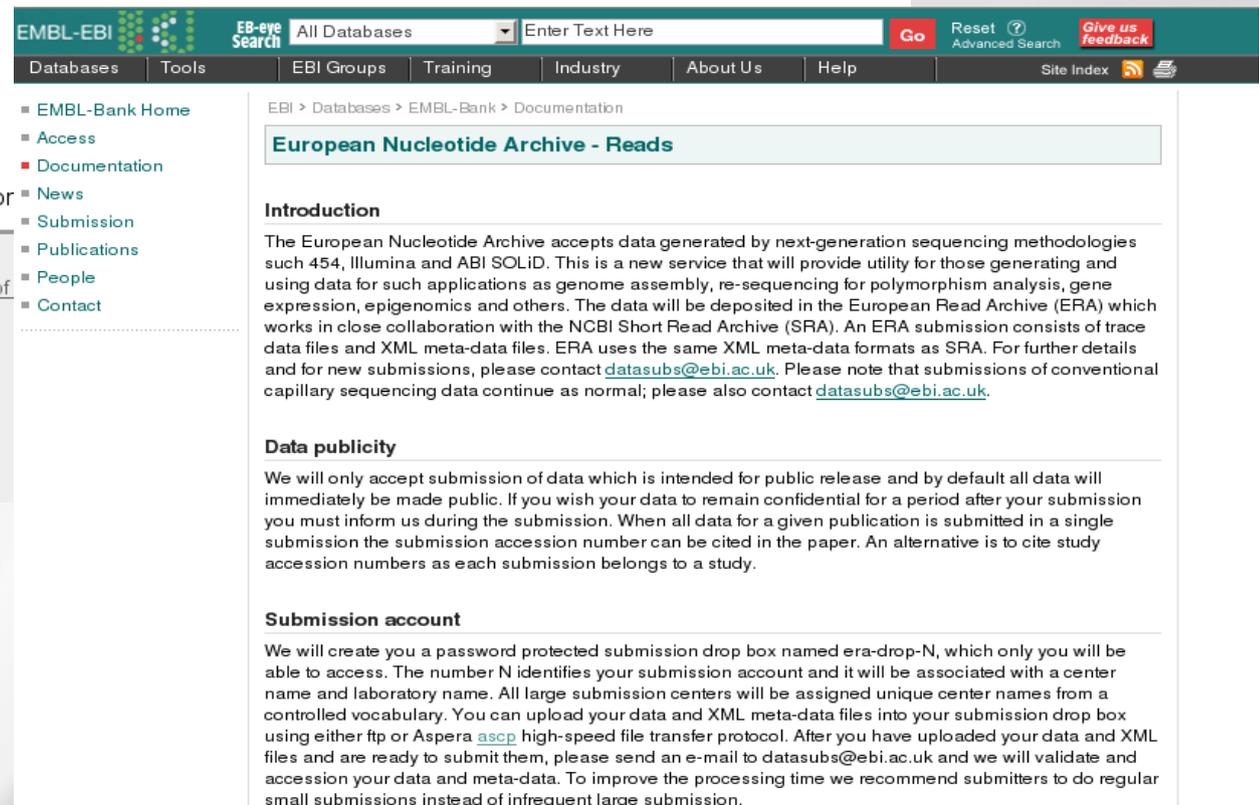
The Sequence Read Archive (SRA) stores raw sequencing data from the "next" generation of sequencing platforms including Roche 454 G...

Current capabilities include:

- [Run Browser](#)
- [Study/Sample/Experiment/Analysis](#) browsers
- [Download facility](#)
- [Search SRA \(using Entrez\)](#)
- [Interactive submissions facility](#)
- [Automated submissions](#)

See [Sequence Read Archive Overview](#) for more information

[Write to the Help Desk](#) | [Privacy Notice](#) | [Disclaimer](#) | [Accessibility](#)
National Center for Biotechnology Information | U.S. National Library of

EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

EMBL-Bank Home
Access
Documentation
News
Submission
Publications
People
Contact

EBI > Databases > EMBL-Bank > Documentation

European Nucleotide Archive - Reads

Introduction

The European Nucleotide Archive accepts data generated by next-generation sequencing methodologies such as 454, Illumina and ABI SOLiD. This is a new service that will provide utility for those generating and using data for such applications as genome assembly, re-sequencing for polymorphism analysis, gene expression, epigenomics and others. The data will be deposited in the European Read Archive (ERA) which works in close collaboration with the NCBI Short Read Archive (SRA). An ERA submission consists of trace data files and XML meta-data files. ERA uses the same XML meta-data formats as SRA. For further details and for new submissions, please contact datasubs@ebi.ac.uk. Please note that submissions of conventional capillary sequencing data continue as normal; please also contact datasubs@ebi.ac.uk.

Data publicity

We will only accept submission of data which is intended for public release and by default all data will immediately be made public. If you wish your data to remain confidential for a period after your submission you must inform us during the submission. When all data for a given publication is submitted in a single submission the submission accession number can be cited in the paper. An alternative is to cite study accession numbers as each submission belongs to a study.

Submission account

We will create you a password protected submission drop box named era-drop-N, which only you will be able to access. The number N identifies your submission account and it will be associated with a center name and laboratory name. All large submission centers will be assigned unique center names from a controlled vocabulary. You can upload your data and XML meta-data files into your submission drop box using either ftp or Aspera [ascp](#) high-speed file transfer protocol. After you have uploaded your data and XML files and are ready to submit them, please send an e-mail to datasubs@ebi.ac.uk and we will validate and accession your data and meta-data. To improve the processing time we recommend submitters to do regular small submissions instead of infrequent large submission.

La présentation des analyses

- Assemblage
 - Utilisation de génomes de référence
- Annotation automatique
 - Utilisation de banques de données validées
- Annotation manuelle
- Recherche de variations
- Recherche de répétitions
 - Utilisation de banques de répétitions

Alignements de séquences

- Données brutes stockées en fichiers indexés
- Différents alignements :
 - Alignement simples ou multiples
 - Alignement nucléique-nucléique, protéique-protéique, nucléique-protéique,...
 - Alignement épissé ou non
 - ...



BLAST

Basic Local Alignment Search Tool

Home

Recent Results

Saved Strategies

Help

➤ Sortie de blast (web)

Mouse-over to show define and scores. Click to show alignments

Color Key for Alignment Scores

<40	40-50	50-80	80-200	>=200
-----	-------	-------	--------	-------

1c111_

Sequences producing significant alignments:	Score	E
	(bits)	Value
refINC_006090.2INC_006090 Gallus gallus chromosome 3, refer...	1253	0.0
refINC_006127.2INC_006127 Gallus gallus chromosome 2, refer...	593	e-167
refINC_006088.2INC_006088 Gallus gallus chromosome 1, refer...	426	e-117
refINC_006089.2INC_006089 Gallus gallus chromosome 2, refer...	422	e-116
refINC_006092.2INC_006092 Gallus gallus chromosome 5, refer...	418	e-114
refINC_006126.2INC_006126 Gallus gallus chromosome W, refer...	410	e-112
refINC_006091.2INC_006091 Gallus gallus chromosome 4, refer...	394	e-107
refINC_006114.2INC_006114 Gallus gallus chromosome 27, refe...	331	3e-88
refINC_006105.2INC_006105 Gallus gallus chromosome 18, refe...	238	1e-59
refINC_006094.2INC_006094 Gallus gallus chromosome 7, refer...	226	1e-56
refINC_006097.2INC_006097 Gallus gallus chromosome 10, refe...	210	6e-52
refINC_006095.2INC_006095 Gallus gallus chromosome 8, refer...	202	2e-49
refINC_006100.2INC_006100 Gallus gallus chromosome 13, refe...	161	5e-37
refINC_006107.2INC_006107 Gallus gallus chromosome 20, refe...	143	1e-31
refINC_006093.2INC_006093 Gallus gallus chromosome 6, refer...	143	1e-31
refINC_006099.2INC_006099 Gallus gallus chromosome 12, refe...	135	3e-29
refINC_006101.2INC_006101 Gallus gallus chromosome 14, refe...	48	0.005
refINC_006102.2INC_006102 Gallus gallus chromosome 15, refe...	48	0.005
refINC_006104.2INC_006104 Gallus gallus chromosome 17, refe...	40	1.3

>[refINC_006090.2INC_006090](#) Gallus gallus chromosome 3, reference assembly (based on Gallus_gallus-2.1)
 Length = 113657789

Score = 1253 bits (632), Expect = 0.0
 Identities = 659/669 (98%), Gaps = 1/669 (0%)
 Strand = Plus / Plus

```

Query: 659      agctgcaggttagaatgaaacctaaacctggtcaaaacggtgggagagacccaaaataca 718
              |||
Sbjct: 113612785 agctgcaggttagaatgaaacctaaacctggtcaaaacggtgggagagacccaaaataca 113612844

Query: 719      acataaaaggaataaaatttgagctaccgagnnnnnnntgaaggaagcacagaagtgg 778
              |||
Sbjct: 113612845 acataaaaggaataaaatttgagctaccgagaaaaaaa-tgaaggaagcacagaagtgg 113612903

Query: 779      aagcagccctggctggaattcgatatgctgcgagagtagcagatacttcaaaaatagaggaa 838
              |||
Sbjct: 113612904 aagcagccctggctggaattcgatatgctgcgagagtagcagatacttcaaaaatagaggaa 113612963
    
```

Terminé

Environnement de visualisation

➤ Vue linéaire du génome

Ensembl Home > Human (GRCh37) Location: 6:133,017,695-133,161,157

Chromosome 6: 133,017,695-133,161,157

Location-based displays: Whole genome, Chromosome summary, Region overview, **Region in detail**, Comparative Genomics, Alignments (image) (51), Alignments (text) (51), Multi-species view (47), Synteny (12), Genetic Variation, Resequencing (2), Linkage Data, Markers, Other genome browsers, UCSC, NCBI

Assembly exception... chromosome 6

Assembly exception... HSCHR6_MHC_AFD, HSCHR6_MHC_COX, HSCHR6_MHC_DBB, HSCHR6_MHC_MAHN, HSCHR6_MHC_MCF, HSCHR6_MHC_OBL, HSCHR6_MHC_SST0

Region overview **Region in detail help** Alignments (Image)

Chromosome bands, Contigs, Ensembl/Havana g... AL023578.1, AL357034.18, AL513524.8, AL032821.2, AL117783.12, AL119102.13, AL591115.6, AL121959.15

ncRNA gene: MOXD1, STX7, TAAR9, TAAR6, TAAR3, TAAR1, TAAR8, TAAR5, TAAR2, VNN1, VNN2, VNN3, C6orf192, RPS12, SNORA101, SNORD100, SNORA33

Gene Legend: Ensembl Homo sapiens version 56.37a (GRCh37) Chromosome 6: 132,589,427 - 133,589,426. Known protein coding, Novel processed transcript, Known RNA gene, Novel pseudogene, Known pseudogene

Location: 6 : 13301769 - 13316115 Go>

Chromosome bands, Human RefSeq/EM... CCDS set, ncRNA gene, Ensembl/Havana g... SNORA33, SNORD100, SNORD101, RPS12-201, RPS12-002

Environnement de visualisation

➤ Carte d'identité d'un transcript

Location: 7:114,055,378-114,330,012 Gene: FOXP2 Transcript: FOXP2-020

Transcript: FOXP2-020 (ENST00000393489)

Forkhead box protein P2 (CAG repeat protein 44)(Trinucleotide repeat-containing gene 10 protein) [Source:UniProtKB/Swiss-Prot;Acc:O15409]

Location [Chromosome 7: 114,055,378-114,330,012](#) forward strand.

Gene This transcript is a product of gene [ENSG00000128573](#) - There are 23 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
FOXP2-001	ENST00000350908	ENSP00000265436	protein_coding
FOXP2-002	ENST00000440349	ENSP00000395552	nonsense_mediated_decay
FOXP2-003	ENST00000324462	ENSP00000319424	protein_coding
FOXP2-004	ENST00000408937	ENSP00000386200	protein_coding
FOXP2-005	ENST00000393498	ENSP00000377135	protein_coding
FOXP2-006	ENST00000412402	ENSP00000405470	nonsense_mediated_decay
FOXP2-007	ENST00000462331	ENSP00000418100	protein_coding
FOXP2-008	ENST00000441290	ENSP00000416825	nonsense_mediated_decay
FOXP2-009	ENST00000459666	No protein product	processed_transcript
FOXP2-010	ENST00000378237	ENSP00000367482	protein_coding
FOXP2-011	ENST00000495516	No protein product	processed_transcript
FOXP2-014	ENST00000393494	ENSP00000377132	protein_coding
FOXP2-015	ENST00000452963	ENSP00000409826	protein_coding
FOXP2-016	ENST00000390668	ENSP00000375084	protein_coding
FOXP2-017	ENST00000360232	ENSP00000353367	protein_coding
FOXP2-018	ENST00000393495	ENSP00000377133	protein_coding
FOXP2-019	ENST00000484307	ENSP00000420262	protein_coding
FOXP2-020	ENST00000393489	ENSP00000377129	protein_coding
FOXP2-201	ENST00000393491	ENSP00000377130	protein_coding
FOXP2-202	ENST00000393500	ENSP00000377137	protein_coding
FOXP2-203	ENST00000403100	ENSP00000384090	protein_coding
FOXP2-204	ENST00000403559	ENSP00000385069	protein_coding
FOXP2-205	ENST00000403611	ENSP00000385497	protein_coding

Transcript and Gene level displays

In Ensembl a gene is made up of one or more transcripts. Views in Ensembl are separated into Gene based views and Transcript based views according to which level the information is more appropriately associated. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

Transcript summary [help](#)



Statistics Exons: 18 Transcript length: 2,285 bps Translation length: 623 residues

CCDS This transcript is a member of the Human CCDS set: [CCDS5761](#)

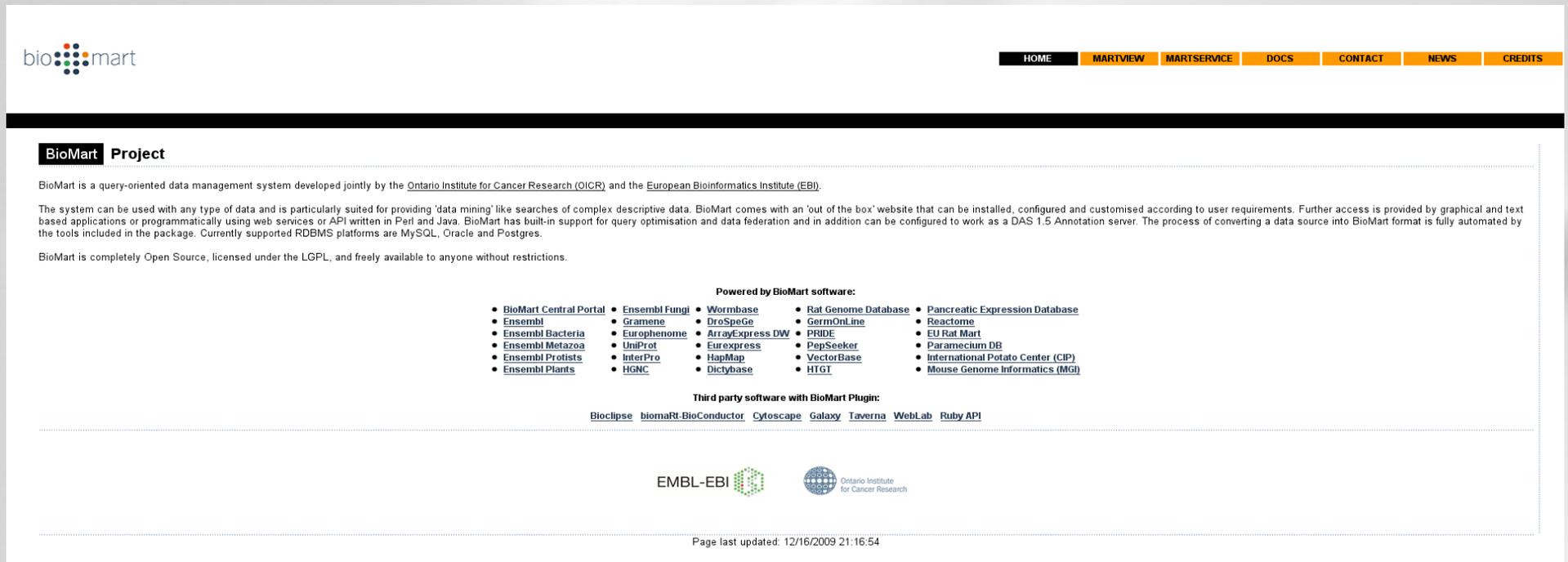
Type Known protein coding

Prediction Method Transcript where the Ensembl genebuild transcript and the [Vega](#) manual annotation have the same sequence, for every base pair. See [article](#).

Alternative transcripts This Ensembl/Havana merge gene entry corresponds to the following database identifiers:
Transcript having exact match between ENSEMBL and HAVANA: [OTTHUMT00000354233](#) [\[view all locations\]](#)

Environnement d'interrogation

- BioMart (www.biomart.org)



The screenshot shows the BioMart website interface. At the top left is the BioMart logo. A navigation bar contains links for HOME, MARTVIEW, MARTSERVICE, DOCS, CONTACT, NEWS, and CREDITS. The main content area is titled "BioMart Project" and includes a description of the system, its capabilities, and licensing. It lists various data sources powered by BioMart software and third-party software with BioMart plugins. The footer includes logos for EMBL-EBI and the Ontario Institute for Cancer Research, along with a page update timestamp.

BioMart Project

BioMart is a query-oriented data management system developed jointly by the [Ontario Institute for Cancer Research \(OICR\)](#) and the [European Bioinformatics Institute \(EBI\)](#).

The system can be used with any type of data and is particularly suited for providing 'data mining' like searches of complex descriptive data. BioMart comes with an 'out of the box' website that can be installed, configured and customised according to user requirements. Further access is provided by graphical and text based applications or programmatically using web services or API written in Perl and Java. BioMart has built-in support for query optimisation and data federation and in addition can be configured to work as a DAS 1.5 Annotation server. The process of converting a data source into BioMart format is fully automated by the tools included in the package. Currently supported RDBMS platforms are MySQL, Oracle and Postgres.

BioMart is completely Open Source, licensed under the LGPL, and freely available to anyone without restrictions.

Powered by BioMart software:

- [BioMart Central Portal](#)
- [Ensembl](#)
- [Ensembl Bacteria](#)
- [Ensembl Metazoa](#)
- [Ensembl Protists](#)
- [Ensembl Plants](#)
- [Ensembl Fungi](#)
- [Gramene](#)
- [Europhenome](#)
- [UniProt](#)
- [InterPro](#)
- [HGNC](#)
- [Wormbase](#)
- [DroSpeGe](#)
- [ArrayExpress DW](#)
- [Eurexpress](#)
- [HapMap](#)
- [Dictybase](#)
- [Rat Genome Database](#)
- [GermOnLine](#)
- [PRIDE](#)
- [PepSeeker](#)
- [VectorBase](#)
- [HTGT](#)
- [Pancreatic Expression Database](#)
- [Reactome](#)
- [EU Rat Mart](#)
- [Paramecium DB](#)
- [International Potato Center \(CIP\)](#)
- [Mouse Genome Informatics \(MG\)](#)

Third party software with BioMart Plugin:

[Bioclipse](#) [biomaRT-BioConductor](#) [Cytoscape](#) [Galaxy](#) [Taverna](#) [WebLab](#) [Ruby API](#)

EMBL-EBI   Ontario Institute for Cancer Research

Page last updated: 12/16/2009 21:16:54

Environnement d'interrogation

- Étape 1 : sélection du lot de données

The screenshot shows the bioMart web interface. At the top left is the 'bio·mart' logo. On the top right, there is a navigation bar with links: HOME, MARTVIEW, MARTSERVICE, DOCS, CONTACT, NEWS, and CREDITS. Below the navigation bar is a toolbar with buttons for 'New', 'Count', and 'Results'. On the far right of the toolbar are icons for 'URL', 'XML', 'Perl', and 'Help'. The main content area is divided into two sections. The top section has a 'Dataset' dropdown menu set to 'ENSEMBL 56 GENES (SANGER UK)'. Below it is a 'Filters' section with a dropdown menu set to 'Gallus gallus genes (WASHUC2)'. Underneath the filters, there are two 'Attributes' listed: 'Ensembl Gene ID' and 'Ensembl Transcript ID'. The bottom section is another 'Dataset' dropdown menu, currently set to '[None Selected]'.

bio·mart

HOME MARTVIEW MARTSERVICE DOCS CONTACT NEWS CREDITS

New Count Results URL XML Perl Help

Dataset: ENSEMBL 56 GENES (SANGER UK)

Gallus gallus genes (WASHUC2)

Filters: Gallus gallus genes (WASHUC2)

Attributes: Ensembl Gene ID, Ensembl Transcript ID

Dataset: [None Selected]

Environnement d'interrogation

➤ Étape 2 : filtres

bio::mart

HOME MARTVIEW MARTSERVICE DOCS CONTACT NEWS CREDITS

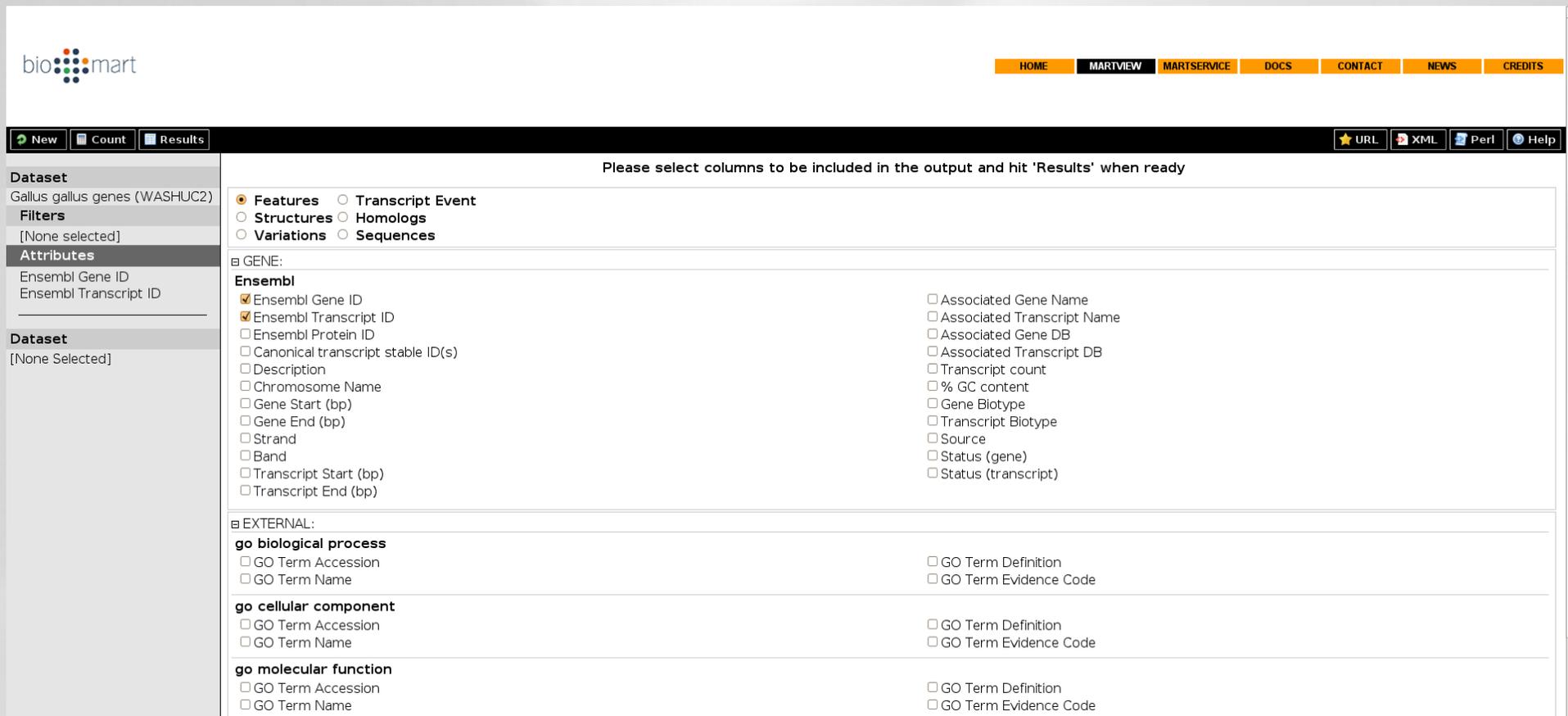
New Count Results URL XML Perl Help

Please restrict your query using criteria below

Dataset Gallus gallus genes (WASHUC2)	<input type="checkbox"/> REGION:
Filters [None selected]	<input type="checkbox"/> GENE:
Attributes Ensembl Gene ID Ensembl Transcript ID	<input type="checkbox"/> Limit to genes ... with WikiGene ID(s) <input checked="" type="radio"/> Only <input type="radio"/> Excluded
Dataset [None Selected]	<input type="checkbox"/> ID list limit Ensembl Gene ID(s) <input type="text" value=""/> <input type="button" value="Parcourir..."/>
	<input type="checkbox"/> Transcript count >= <input type="text" value=""/>
	<input type="checkbox"/> Gene type miRNA misc_RNA Mt_rRNA Mt_tRNA protein_coding
	<input type="checkbox"/> Source ensembl
	<input type="checkbox"/> Status (gene) KNOWN
	<input type="checkbox"/> Status (transcript) KNOWN
	<input type="checkbox"/> GENE ONTOLOGY:
	<input type="checkbox"/> MULTI SPECIES COMPARISONS:
	<input type="checkbox"/> PROTEIN DOMAINS:
	<input type="checkbox"/> VARIATIONS:

Environnement d'interrogation

➤ Étape 3 : attributs



The screenshot shows the bio.mart web interface. At the top left is the bio.mart logo. A navigation bar contains links for HOME, MARTVIEW, MARTSERVICE, DOCS, CONTACT, NEWS, and CREDITS. Below the navigation bar is a toolbar with buttons for New, Count, Results, URL, XML, Perl, and Help. The main content area is titled "Please select columns to be included in the output and hit 'Results' when ready".

On the left side, there is a sidebar with the following sections:

- Dataset:** Gallus gallus genes (WASHUC2)
- Filters:** [None selected]
- Attributes:** Ensembl Gene ID, Ensembl Transcript ID
- Dataset:** [None Selected]

The main content area contains the following options for selection:

- Features** (selected) / Transcript Event
- Structures / Homologs
- Variations / Sequences

Below these are expandable sections:

- GENE:**
 - Ensembl**
 - Ensembl Gene ID
 - Ensembl Transcript ID
 - Ensembl Protein ID
 - Canonical transcript stable ID(s)
 - Description
 - Chromosome Name
 - Gene Start (bp)
 - Gene End (bp)
 - Strand
 - Band
 - Transcript Start (bp)
 - Transcript End (bp)
 - Associated Gene Name
 - Associated Transcript Name
 - Associated Gene DB
 - Associated Transcript DB
 - Transcript count
 - % GC content
 - Gene Biotype
 - Transcript Biotype
 - Source
 - Status (gene)
 - Status (transcript)
- EXTERNAL:**
 - go biological process**
 - GO Term Accession
 - GO Term Name
 - GO Term Definition
 - GO Term Evidence Code
 - go cellular component**
 - GO Term Accession
 - GO Term Name
 - GO Term Definition
 - GO Term Evidence Code
 - go molecular function**
 - GO Term Accession
 - GO Term Name
 - GO Term Definition
 - GO Term Evidence Code

Conclusions

- Beaucoup de fichiers indexés (données brutes), des bases de données relationnelles pour les données élaborées.
- La quantité de données générées augmente très vite.
- Le besoin en moyens de calcul et de stockage est important.
- **Notre connaissance du fonctionnement des génomes est encore très limitée.**