

# Analyse de Séquences

M1 BIBS

2<sup>e</sup> partie

# Les programmes Génome

# Les programmes Génome

- **Préhistoire**

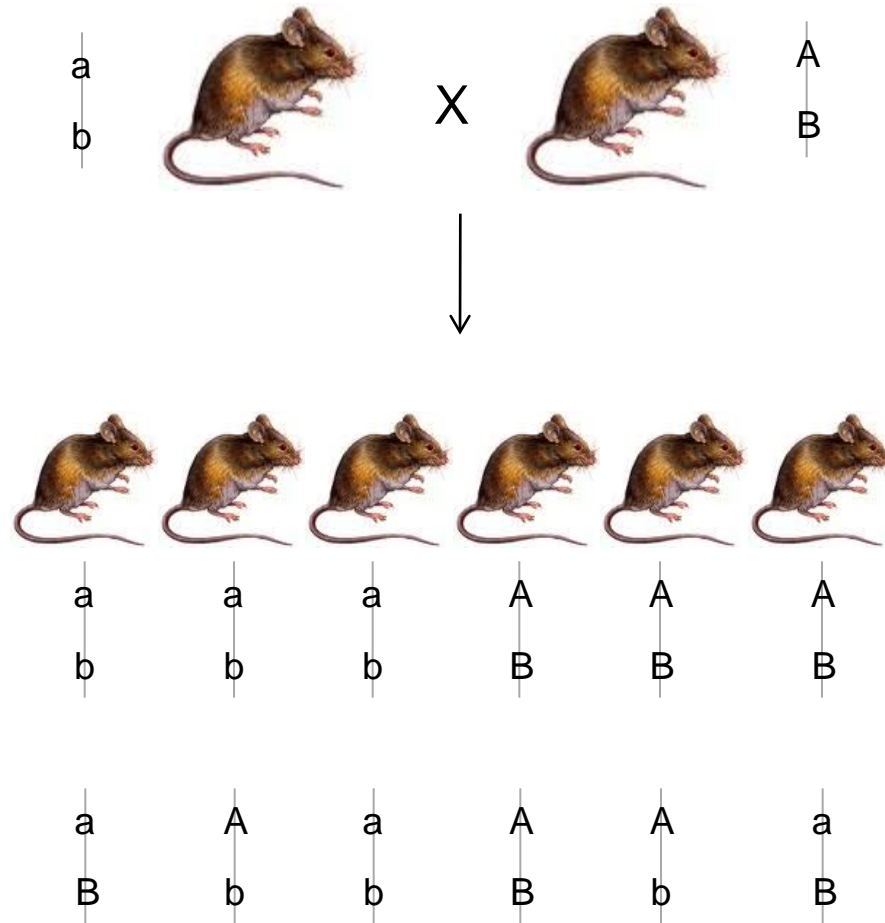
- Séquençage de Sanger
- Avant le séquençage d'organismes: séquençage de bactériophages.
- Plus grande séquence obtenue avant projets « Génome »: bactériophage 1: 50.000pb
- Comment passer à plusieurs millions de bp?

<b>Organisme</b>	<b>Bp</b>	<b>Genes</b>	<b>ratio</b>
HIV-1	10 000	10	1000
<i>Haemophilus influenzae</i>	1 830 000	1703	1075
<i>Escherichia coli</i>	4 600 000	4288	1072
<i>Methanococcus jannashchii</i>	1 660 000	1738	955
<i>Amoeba dubia</i>	670 000 000 000	~5000?	134 000 000
<i>Amoeba proteus</i>	270 000 000 000	~5000?	54 000 000
<i>Saccharomyces cerevisiae</i>	13 000 000	5885	2209
<i>Erysiphe cichoracearum</i> (champignon)	1 500 000 000	~10 000?	150 000
<i>Coscinodiscus asteromphalus</i> (diatomée)	25 000 000 000	~5000?	5 000 000
<i>Caenorhabditis elegans</i>	100 000 000	~14 000	7000
<i>Drosophila melanogaster</i>	170 000 000	~12 000	14 000
<i>Arabidopsis thaliana</i>	120 000 000	~10 000	12 000
<i>Lilium formosanum</i> (nénuphar)	36 000 000 000	~15 000?	2 400 000
<i>Zea mays</i>	2 300 000 000	32 000	70 000
<i>Allium cepa</i> (oignon)	18 000 000 000	~20 000?	900 000
<i>Protopterus aethiopicus</i> (dipneuste)	140 000 000 000	~40 000?	3 500 000
<i>Homo sapiens</i>	3 400 000 000	~20 000	100 000

# Le Programme Génome Humain

- **Le programme Génome tel que défini par le DOE en 1990**
  - Carte génétique, résolution 2 à 5 cM (4 à 10 Mb chez l'homme)
  - Carte physique, résolution 100 kb
  - A long terme, séquençage du génome humain (après amélioration des techniques)
  - Séquençage du génome d'espèces modèles
  - Ajouté en 93: identification du plus grand nombre de gènes possible (par séquençage de cDNA)

# Cartographie génétique: principe



★ Gènes proches

★ Gènes distants

# Cartographie Physique

BUT: rendre disponible toute région correspondant à un marqueur génétique sous forme d'un fragment d'ADN cloné.

- **Banques Génomiques**

- Insérer des fragments de l'ADN à séquencer dans un vecteur (plasmide, phage)
- Multiplier ce vecteur dans un organisme (bactérie, levure) = « cloner »
- Une banque génomique humaine se présente sous la forme d'une collection de plusieurs centaines de microplaques à 96 puits, chacun contenant un clone.

- **Ordonnancement des clones**

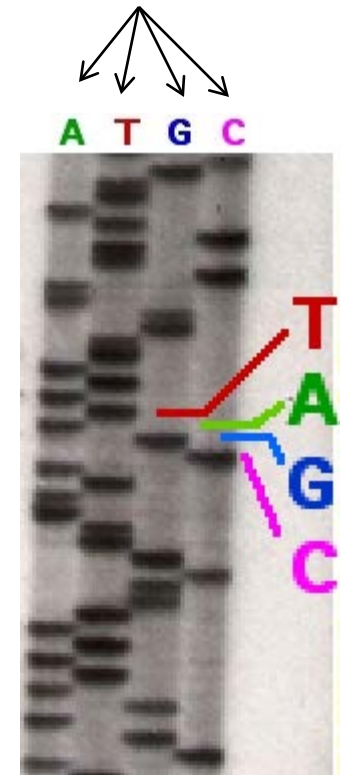
- Si l'on repère un marqueur génétique dans un clone (par exemple par hybridation), ce clone se trouve directement positionné (grâce à la carte génétique).
- On peut également utiliser le recouvrement entre clones pour les positionner les uns par rapport aux autres.

# Le séquençage

# Le séquençage de Sanger (1977)

- Séquençage par terminaison de chaîne
  - Utilisation de dideoxynucléotides pour interrompre la synthèse à un certain type de base.
  - 4 réactions + marquage radioactif
- Amélioré en 1987 par l'introduction de marqueurs fluorescents (1 seule réaction) et l'automatisation.

dideoxynucléotides



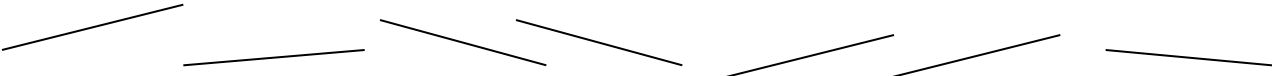
Wikipedia

# Stratégie de séquençage Shotgun (peut se faire avec la méthode Sanger)

Clone à séquencer

---

Fractionnement



Séquençage aléatoire



Assemblage



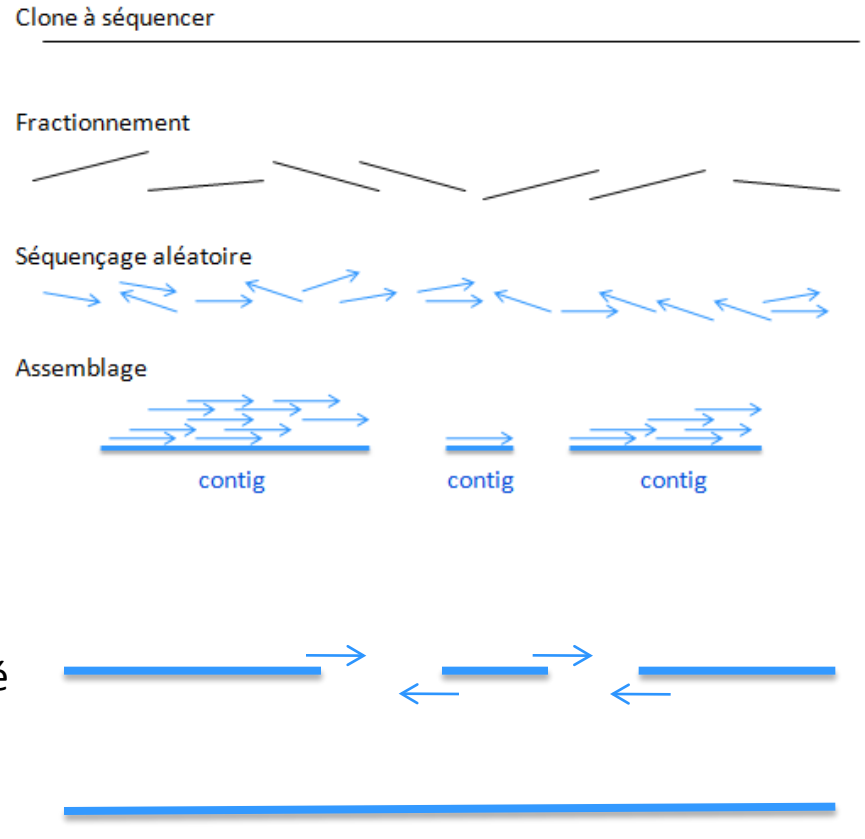
Problématique  
en présence de  
longues régions  
répétées)

Scaffold (si information cartographique)



# Stratégie Shotgun partiel

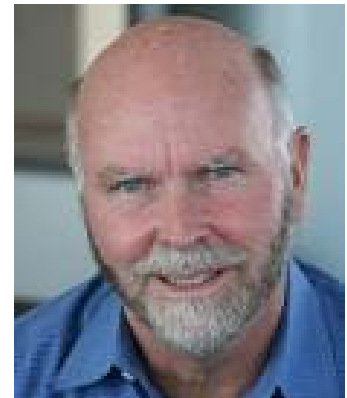
- Séquençage shotgun de clones prélevés aléatoirement
- Les séquences restantes sont réalisées de façon dirigée.



# Shotgun « génome complet »

- le shotgun "genome complet" (whole-genome shotgun) s'applique normalement aux génomes relativement simples, en exploitant au maximum les informations de cartographie et la bioinformatique pour éviter les mésassemblages.
- Employé par la Sté Celera pour le séquençage du génome de la Drosophile, puis du génome humain

Craig Venter



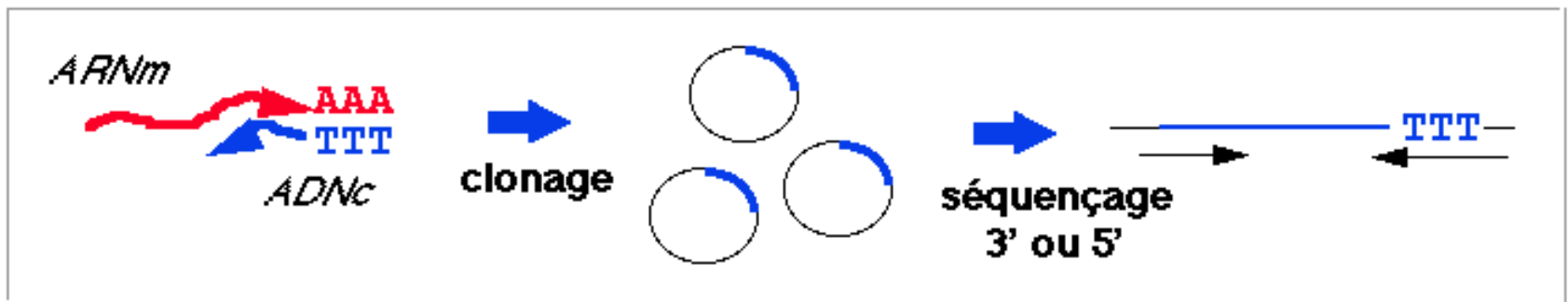
# Les technologies de génomique à haut-débit

- Depuis les années 1990
  - Séquençage massif de cDNA
  - Puces à ADN

# Le séquençage direct des gènes par les cDNA

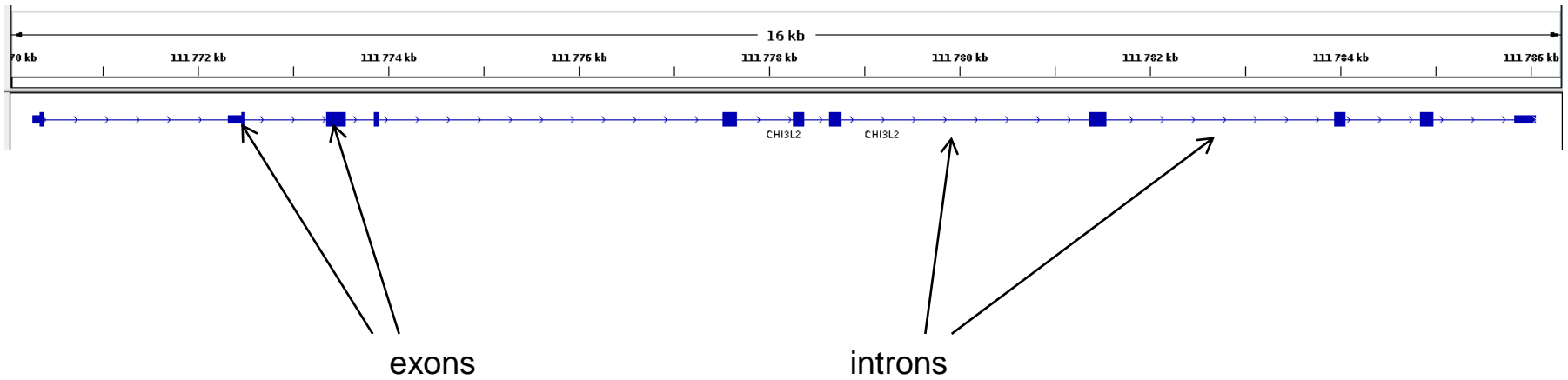
# Les cDNA

- Idée initiale:
  - pourquoi vouloir tout séquencer (95% de junk DNA) si ce sont les gènes qui nous intéressent?
- Usage actuel:
  - Analyse du transcriptome

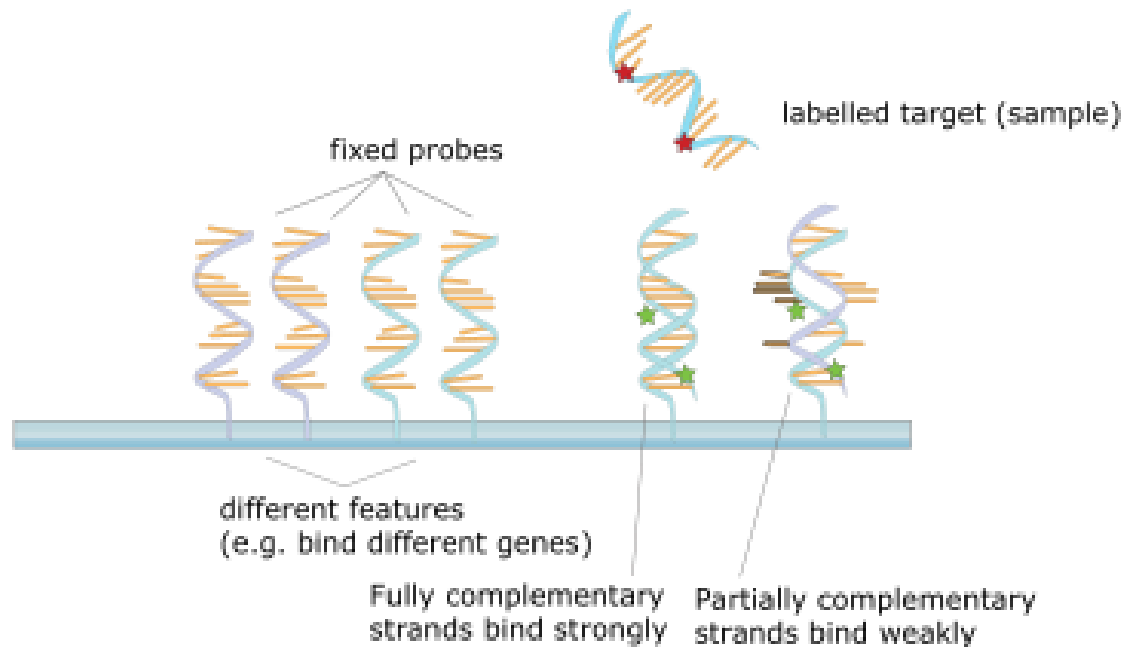


- EST (Expressed sequence Tag) = Séquences partielles d'ADNc clonés et prélevés aléatoirement.
- Full length cDNA: Séquences complète d'ADNc

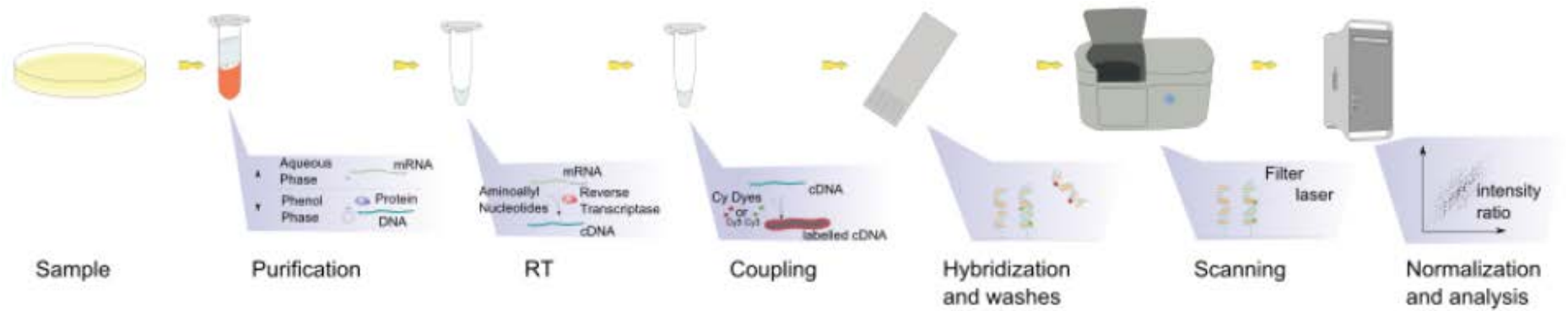
# Le mapping des cDNA sur le génome



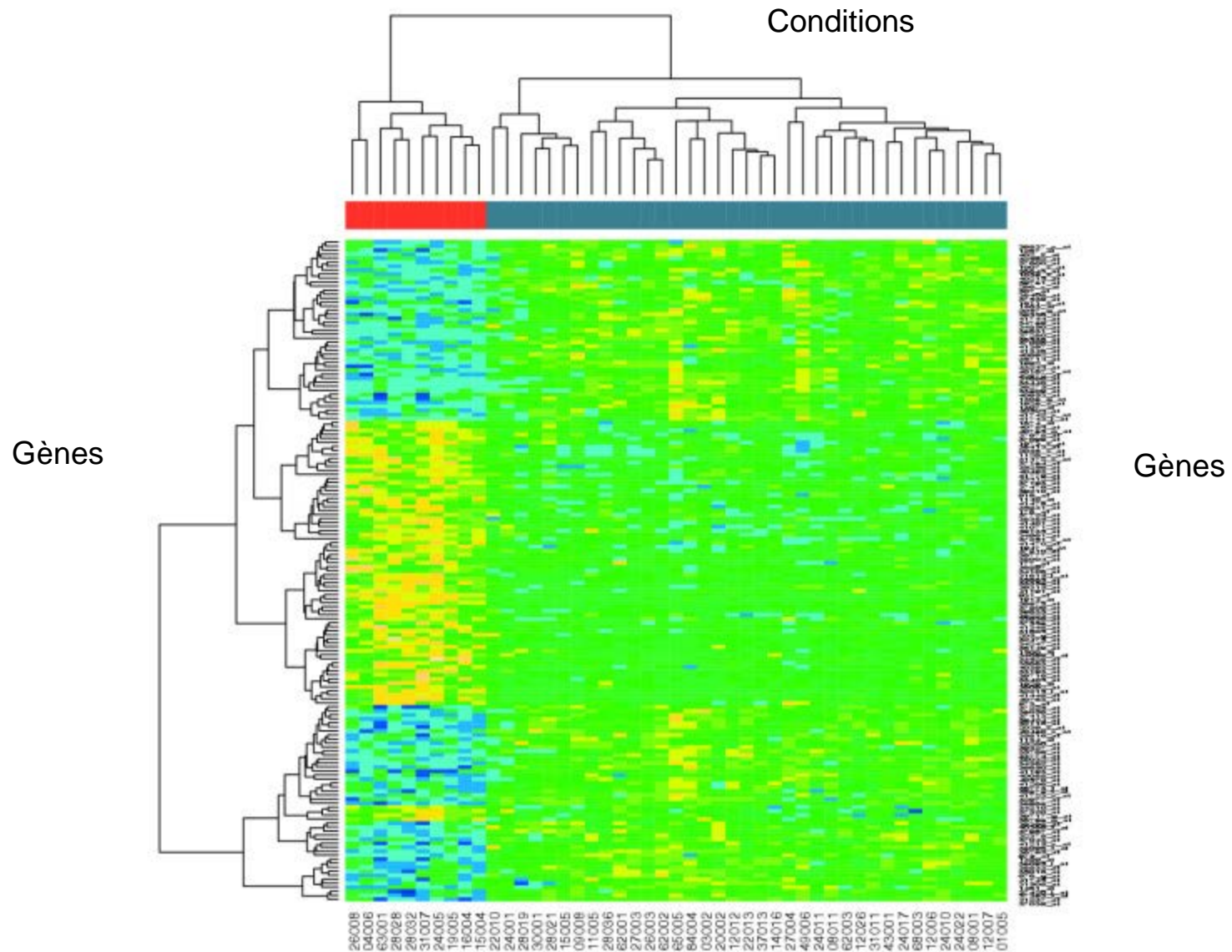
# Les puces à ADN (DNA microarrays)



# Profil d'expression des gènes par puces à ADN



# Profil d'expression des gènes



# Puces à ADN

- De nombreuses variantes
  - CGH array pour comparer l'ADN génomique normal/de tumeur
  - Tiling array pour couvrir l'ensemble du génome
  - Exon array pour rechercher des épissages différentiels
  - CNP array pour rechercher des SNP

# NGS (Next Generation Sequencing)

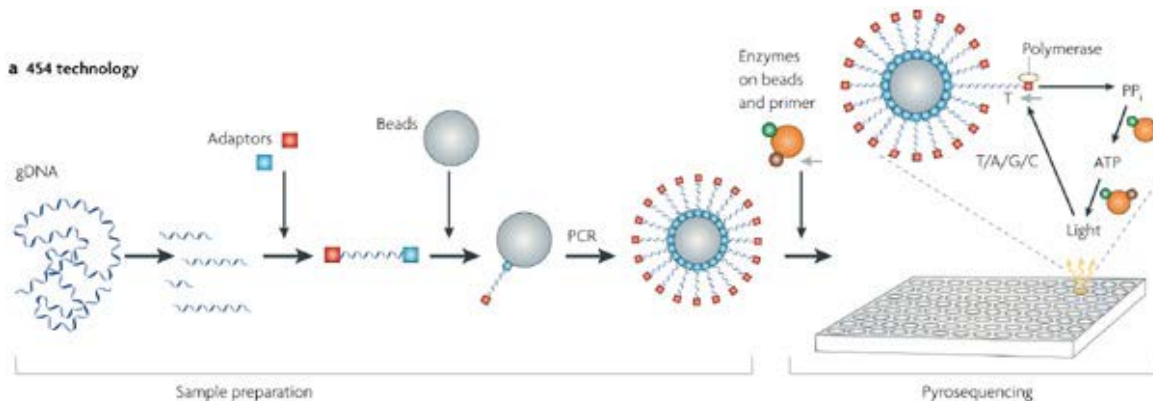
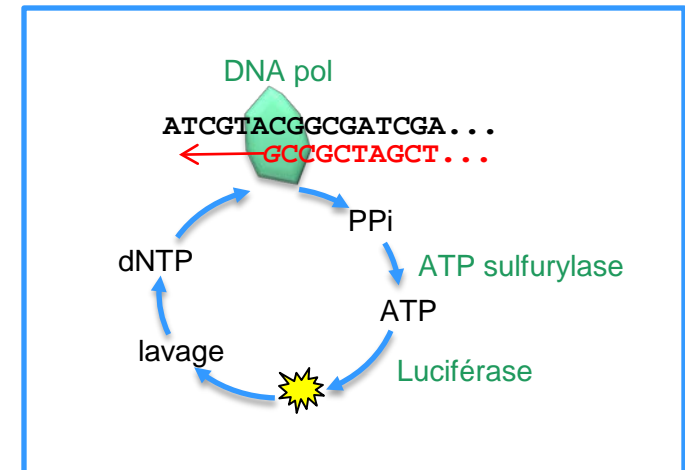
- Depuis 2006

# Le pyroséquençage

454 LifeSciences/Roche  
Biotage/Qiagen

1. ADN immobilisé (billes)
2. Synthèse brin complémentaire par ADN polymérase
3. Introduction des dNTP un par un
4. Si bon dNTP: libération PPI, synthèse ATP, et émission de lumière.

300-500nt à la fois x 400.000

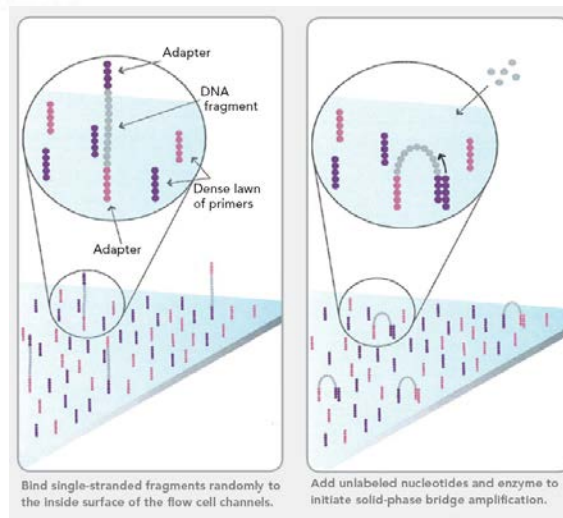
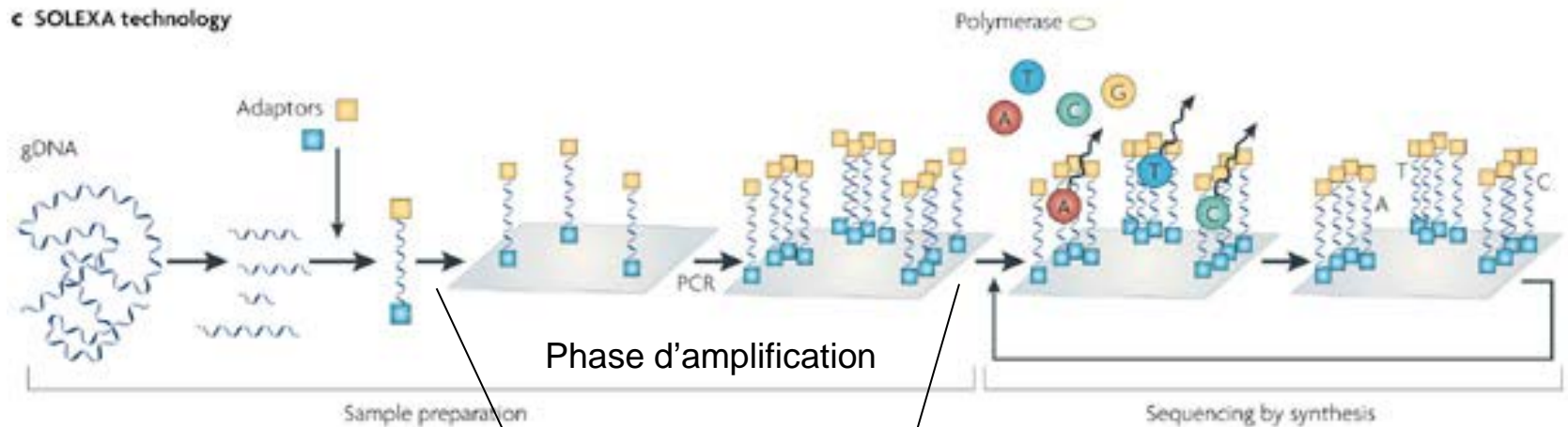


# Autres méthodes de « sequencing by Synthesis »

36-100nt à la fois x 400 millions

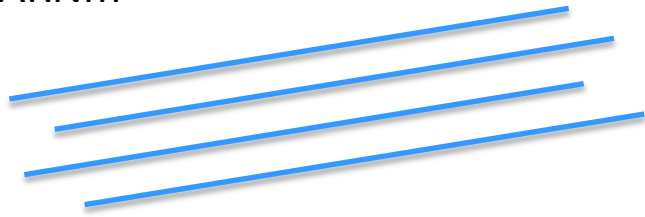
Solexa/Illumina  
Solid

c SOLEXA technology

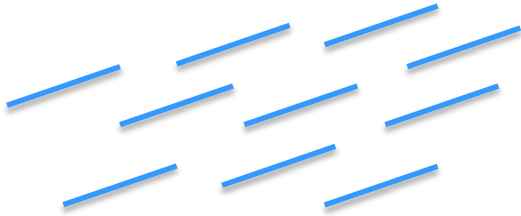


Nature Reviews | Microbiology

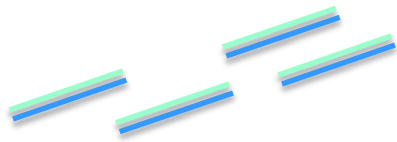
ARNm



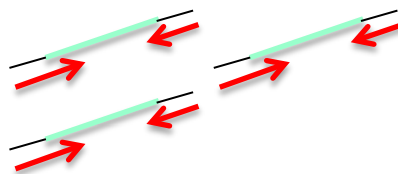
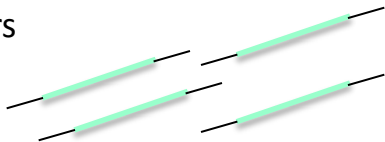
fractionnement



cDNA



adaptateurs



- Transcriptome par séquençage haut-débit.
- On parle aussi de « deep sequencing »
- Peut être précédé d'une étape de filtrage pour petits ARN, permet de pêcher les miRNA, piRNA, etc.

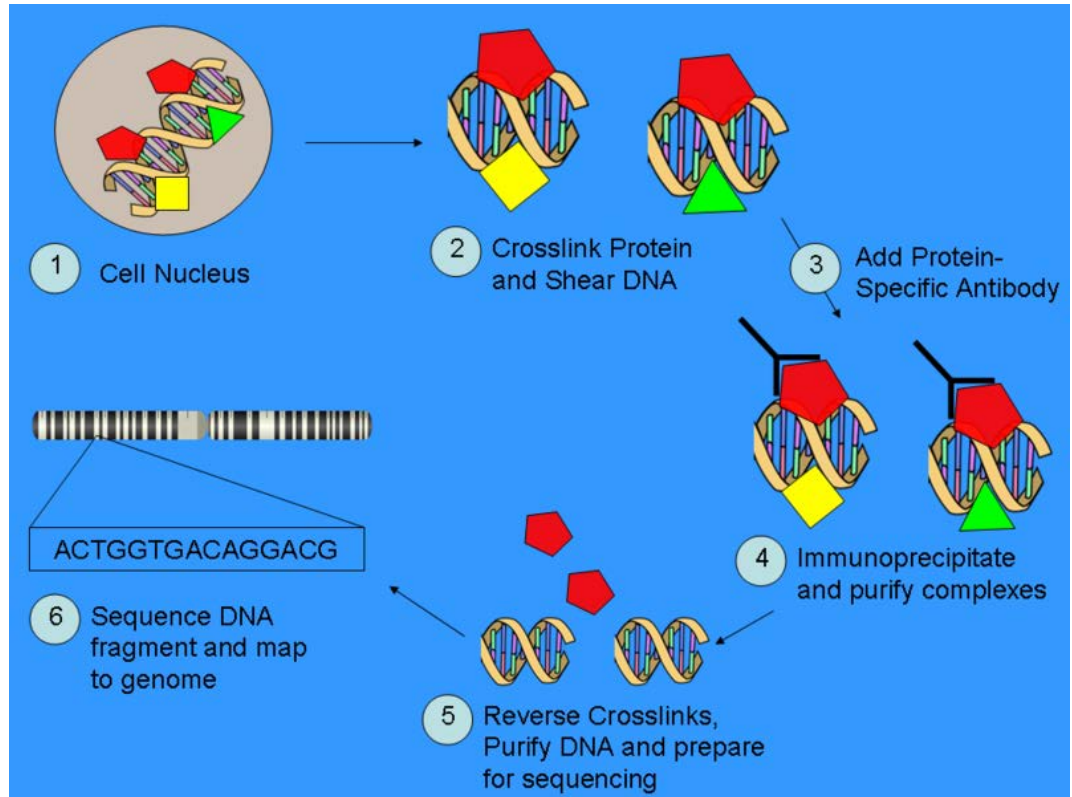
# RNA-Seq



Séquençage

# ChIP-Seq

- ChIP=Chromatin immunoprecipitation
- Permet d'identifier les sites de liaison de protéines (facteurs de transcription, repressseurs, enhancers, etc.) sur l'ADN génomique



Wikipedia

# Metagénomomes

- Fragments d'ADN séquencés aléatoirement à partir d'un environnement donné
- Soit fragments ciblés (p. ex. ARNr 16S), soit ADN génomique
- Environnements étudiés:
  - Océan, intestin humain, drainage minier acide,
- Applications:
  - Populations microbiennes (structure, variations)
  - Découverte et séquençage de microorganismes
  - Découverte de gènes/fonctions

# Vers une utilisation du séquençage pour toutes sortes de questions

- Transcriptome (fin des puces à ADN?)
- Reséquençage, SNP (génomome humain à 700\$)
- Détection d'agents infectieux / bioterrorisme
- ...

# Etat des programmes (2011)

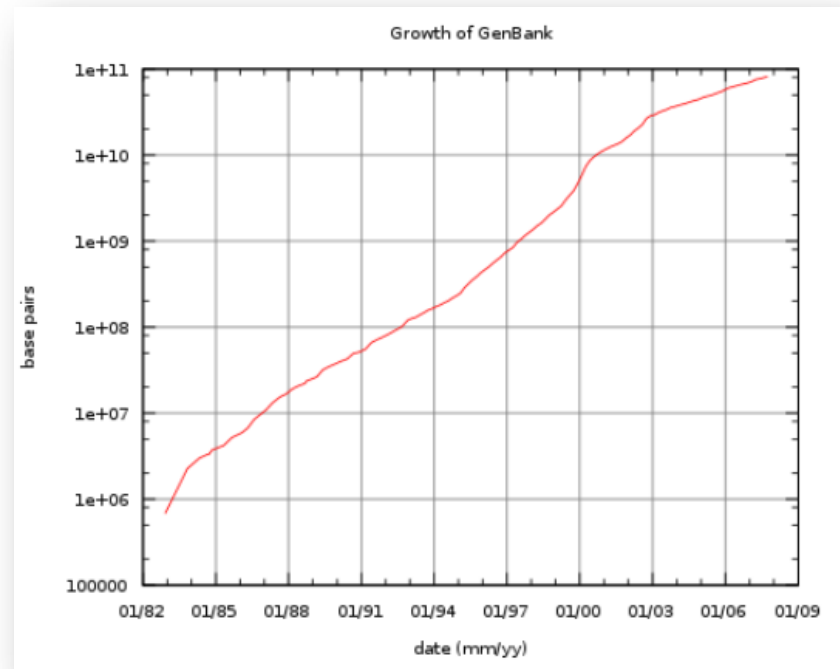
- 1700 génomes procaryotes (110 archae et 1600 bactéries)
  - 1er génome: haemophilus influenza (1995)
- 100+ Génomes eucaryotes:
  - 15 Levures
  - 12 Drosophiles
  - Arabidopsis, riz, mais
  - Humain, souris, rat... >20 mammifères
  - Vertébrés: Fugu, Tetraodon, poulet, chordés, ornithorynque..
  - Protistes
- Milliards de transcrits (pleine longueur et partiels)
- Centaines de métagénomés (Sargasses, GOS, gut, etc...)

# Les banques de données génomiques

# Genbank: La banque d'ADN du NIH

- **Etat Sept 2011**

- $130 \times 10^9$  bases
- $142 \times 10^6$  séquences
- Genbank double environ tous les 14 mois depuis ses débuts en 1982.
- Nouvelle version tous les 2 mois



# Divisions de Genbank

- **ESTs** (Expressed sequence tags):  
Principale division de Genbank. 18 10<sup>6</sup> sequences, 580 organismes différents
- **GSS** (Genome Sequence Survey):  
résultats de séquençages aléatoire de BAC, dans le cadre de projets Génome. Non assemblées, courtes.
- **HTGS** (High Throughput Genomic Sequences): séquences génomiques en cours d'assemblage. Une fois complètement assemblées, les séquences passent dans les divisions « organisme ».
- Bactéries (**BCT**), virus (**VRL**), primates (**PRI**), rongeurs (**ROD**) etc: divisions « organismes ».
- ~20 divisions en tout.

Nb entrées	Nb. bases	Espèce
1355113	854232260	Homo sapiens
378892	179249409	Mus musculus
76471	139699685	Caenorhabditis elegans
66177	69663817	Arabidopsis thaliana
48963	53428355	Drosophila melanogaster
10571	28658828	Saccharomyces cerevisiae
39568	25816686	Rattus norvegicus
4923	17859484	Escherichia coli
32221	16490243	Fugu rubripes
31480	13072925	Oryza sativa
28406	11746328	Rattus sp.
9540	10912762	Schizosaccharomyces pombe
24125	10712174	Human immunodeficiency virus type 1
1086	9893044	Bacillus subtilis
15370	5794059	Brugia malayi
661	5701954	Mycobacterium tuberculosis
4852	5585160	Gallus gallus
4680	5400457	Plasmodium falciparum
5063	4559072	Bos taurus
10845	4409926	Toxoplasma gondii

Organismes dans Genbank (en 2002)

# Enregistrement Genbank

- Chaque enregistrement se voit attribuer un numéro d'accession, stable et unique, et chaque séquence un numéro GI.
- Quand un changement est effectué dans un enregistrement Genbank, le num. d'accession reste, le GI change.

# Enregistrement Genbank avec annotation

```

LOCUS       L10986                47233 bp    DNA        linear   INV 21-SEP-2004
DEFINITION  Caenorhabditis elegans cosmid F10E9, complete sequence.
ACCESSION   L10986
VERSION     L10986.2  GI:38638818
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 47233)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  PUBMED    9851916
FEATURES             Location/Qualifiers
     source          1..47233
                    /organism="Caenorhabditis elegans"
                    /mol_type="genomic DNA"
                    /strain="Bristol N2"
                    /db_xref="taxon:6239"
                    /chromosome="III"
                    /clone="F10E9"
     gene            265..26728
                    /gene="mig-10"
                    /locus_tag="F10E9.6"
     CDS             join(265..338,3266..3515,15194..15317,21507..21
                    21727..21887,23171..23335,24302..24472,24524..24608,
                    25012..25827,26284..26430,26478..26728)
                    /gene="mig-10"
     /translation="MDSCEECDLEVDSDEEDQLFGKECISLLSLLPLSSSTLLSNA
                    INLELDEVERPPPLLNVLVEEQQFPKVCANIEEENELEADTEEDIAETADDEESKDPVE
                    KTFENFEPVMTDITYDFPDYPVQIRARPVPPKPPIDTVRYSMNNIKESADWQLDELL
                    EELEALETQLNSSNGDQLLLGVSGIPASSSRENVKSISTLPPPPALSYHQTPQQPQ
                    . . .
                    QVYTIGIGWEKKYKSPTPWCISIKLTALQMKRSQFIKYICAEDEMTFKKWLVALRIAKN
                    GAELLENYERACQIRRETLGPASSMSAASSSTAISEVPHLSLHHQRTPSVASSIQLSS
                    HMMNNPETHPLSVNVRNQSPASFVNSCQQSHPSRTSAKLEIQYDEQPTGTIKRAPLDV
                    LRRVSRASTSSPTIPQEESDSDEEFAPPPVAVSMRMPPTVTPPKPCTPLTSKKAAPP
                    PPKRSDTTKLQASMPAPAKNDLEAALARREKMATMEC"
  
```

BASE COUNT	2598 a	2024 c	1888 g	2449 t		
ORIGIN						
1	ttctaaaagt	cgaaaaacga	gcaatTTTTg	atgctagatt	tttTgattg	acgaatTTTT
61	tcagTttttt	ttctttaaaa	aaggtTTTTg	accctttaa	gttttcttt	cccttccaat
121	tttttccttc	tttcttatac	gacttctcaa	gtttcaactc	taaaacaag	ctacatgtac
181	atttccggtg	aactttgtgt	ctcagaagat	ccattttctt	ttgtttacat	ttattcaaga
241	ttgaattcca	aaatttcagc	caatatggac	agttgcgaag	aggaatgcca	cttggaagtt
301	gacagtgacg	aagaagatca	actttttggt	gaaaagtggg	gagttcttat	tgtggtaacc
361	aaagaaatgt	cagtggtccg	taaacacttg	actcccaaat	ggtttctcgt	aattacctta
421	tgcacacttt	tcaagtgttt	gccgtttgat	cttagccaat	ttgaaacggt	tagatgttaa
481	atggaaaaatg	ggtaaaagt	tttattttat	agaaaaaagg	ttggaaaaaa	aatcagagtca
541	ctgaatagtt	tgaagaacgg	aaaaataaaa	ctttccaaaa	atcataaaac	atttagtggt
601	tcgaaaatta	tagtgTTTT	ttgtttggtg	tgTTTTgaca	aaagctaaac	catctttatt
661	gtagTtttgt	aaaatgttca	caaagatgcg	tttttttttc	aaatttggca	ggctatcttt
721	acattcacat	ttggataatt	caaatttttc	ttatcgctaa	caaattttcc	tatttttcca
781	attattcggt	tttataaaagc	tttggtagta	tgTgtgtct	atcttttagt	gtcatcagtt
	. . .					
	//					

# Séquence au Format Fasta

```
>U00096 Escherichia coli K-12 MG1655 complete genome.  
agcttttcatttctgactgcaacgggcaatatgtctctgtgtggattaaaaaaagagtgtc  
tgatagcagcttctgaactggttacctgccgtgagtaaattaaaattttattgacttagg  
tcactaaatactttaaccaatataggcatagcgcacagacagataaaaattacagagtac  
acaacatccatgaaacgcattagcaccaccattaccaccaccatcaccattaccacaggt  
aacggtgcgggctgacgcgtacaggaaacacagaaaaaagcccgcacctgacagtgcggg  
ctttttttttcgaccaaaggtaacgaggtaacaaccatgagagtggtgaagttcggcggg  
acatcagtggaatgcagaacgttttctgcgtgttgccgatattctggaaagcaatgcc  
aggcaggggcaggtggccaccgtcctctctgccccgccaaaatcaccaaccacctggtg  
gcatgattgaaaaaacattagcggccaggatgctttaccaatatacagcagtgccgaa  
cgtatttttgccgaacttttgacgggactcgccgccgccagccgggggttcccgtggcg  
caattgaaaactttcgtcgatcaggaatttgcccaataaaacatgtcctgcatggcatt  
agtttggtggggcagtgcccggatagcatcaacgctgcgctgatttgccgtggcgagaaa  
atgtcgatcgccattatggccggcgtattagaagcgcgcggtcacaacgttactgttatc  
gatccggtcgaaaaactgctggcagtggggcattacctcgaatctaccgtcgatattgct  
gagtcacccgccggtattgaggcaagccgcattccggctgatcacatgggtgctgatggca  
ggtttcaccgccggtaatgaaaaaggcgaactgggtgggtgcttggacgcaacggttccgac  
tactctgctgcgggtgctggctgcctgtttacgcgccgattggttgcgagatttggacggac
```

# Autres banques nucléotidiques

- EMBL: Equivalent européen de Genbank. Format différent, contenu presque identique.
- DDBJ: équivalent au Japon
- Banques spécialisées Certaines collections de séquences, bien que généralement présentes dans Genbank, sont beaucoup plus utiles lorsqu'elles sont rassemblées dans des banques spécialisées, par ex:
  - Récepteurs des lymphocytes T (Réarrangements de l'ADN)
  - Génomes HIV, etc.
- Banques pour Blast
  - NR nucléique (« Non-redundant »). All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1 or 2 HTGS sequences). (n'est plus "non-redondant")
  - DbEST: dbest Database of GenBank+EMBL+DDBJ sequences from EST Divisions

# SRA: short read archive

- Part of the « European Nucleotide Archive » (sequences brutes issues de séquenceurs).
- Information sur chaque entrée:
  - Study
  - Sample
  - Experiment
  - Run
  - Organism
  - Instrument Platform
  - Library Name
  - Read Count
  - Base Count
  - File Name / File Size
- Séquences au format fastq

# Format fastq

Descriptif du read (position sur la piste de séquençage, taille,..)



```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36  
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC  
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
```



Qualité (probabilité que la base soit correcte) encodé par code ASCII

Attention: plusieurs versions (Illumina, Sanger..)

# Banques protéiques

- Swissprot (UniProtKB/Swissprot).
  - La mieux annotée des banques protéiques. 2011: 530.000 entrées.
  - Curation par experts seulement (basé sur publis)
  - Attention: toutes les protéines connues n'y sont pas!
    1. Evidence at protein level 72765 13.7%
    2. Evidence at transcript level 69863 13.1%
    3. Inferred from homology 373177 70.1%
    4. Predicted 14474 2.7%
    5. Uncertain 1867 0.4%

# Banques protéiques

- TrEMBL (UniProtKB/TrEMBL):
  - banque protéique produite automatiquement par traduction banque EMBL. 2011: 17.000.000 entrées
- Uniprot=Swissprot+TrEMBL
- Dizaines de Banques spécialisées
  - Cazy (Carbohydrate Active Enzymes)
  - ABC transporters
  - etc.
- Pour Blast
  - NR Protéique (Non-redundent): Banque protéique du NCBI = Traduction de tous les CDS de GenBank + PDB + SwissProt + PIR + PRF - redondances.

# Entrez (recherche multi-bases)

The screenshot shows the Entrez cross-database search interface in a Netscape browser window. The browser title is "Entrez cross-database search - Netscape" and the address bar shows "http://www.ncbi.nlm.nih.gov/Entrez/". The interface features the NCBI logo and the Entrez logo with the tagline "Entrez, The Life Sciences Search Engine". A navigation menu includes "HOME", "SEARCH", "SITE MAP", "PubMed", "Entrez", "Human Genome", "GenBank", "Map Viewer", and "BLAST". A search bar is present with "GO", "CLEAR", and "Help" buttons. The main content area is titled "Welcome to the new Entrez cross-database search page" and lists various databases with icons and descriptions:

- PubMed:** biomedical literature citations and abstracts
- PubMed Central:** free, full text journal articles
- Books:** online books
- OMIM:** online Mendelian Inheritance in Man
- Site Search:** NCBI web and FTP sites
- Nucleotide:** sequence database (GenBank)
- Protein:** sequence database
- UniGene:** gene-oriented clusters of transcript sequences
- CDD:** conserved protein domain database
- Genome:** whole genome sequences
- 3D Domains:** domains from Entrez Structure
- Structure:** three-dimensional macromolecular structures
- UniSTS:** markers and mapping data
- Taxonomy:** organisms in GenBank
- PopSet:** population study data sets
- SNP:** single nucleotide polymorphism
- GEO Profiles:** expression and molecular abundance profiles
- Gene:** gene-centered information
- GEO DataSets:** experimental sets of GEO data
- HomoloGene:** eukaryotic homology groups
- Cancer Chromosomes:** cytogenetic databases
- PubChem Compound:** small molecule chemical structures
- PubChem BioAssay:** bioactivity screens of chemical substances
- PubChem Substance:** chemical substances screened for bioactivity
- Journals:** detailed information about the journals indexed in PubMed and other Entrez databases
- MeSH:** detailed information about NLM's controlled vocabulary
- NLM Catalog:** catalog of books, journals, and audiovisuals in the NLM collections

# Ensembl ([www.ensembl.org](http://www.ensembl.org))

- Plusieurs banques en une:
  - Génomes assemblés
  - Peptides confirmés
  - Transcrits confirmés
  - peptides prédits
  - Transcrits prédits
- + un genome browser

- Méthode de prédiction (système Genewise): Utilisent programme de prédiction Genscan, puis Blast contre: protéines, mRNA, EST
- Génome humain version Sep 2008 (NCBI 36) : Confirmed protein-coding genes: 21649; RNA genes: 4810; Predicted genes (Genscan): 49796; base pairs:  $3,25 \cdot 10^9$ .

Species - Ensembl v24		
Human	pre!	NCBI 34 Jul 04
Mouse		NCBI m33 Jul 04
Zebrafish		WTSI Zv4 Sep 04
Rat		RGSC 3.1 Jul 04
Chicken		WASHUC1 Jul 04
Mosquito		MOZ 2 Apr 04
Fugu		Fugu v2.0 May 04
Fruitfly		BDGP 3.1 Jul 03
Chimp		CHIMP1 May 04
Honeybee		Amel1.1 Sep 04
Tetraodon		TETRAODON7 Sep 04
Dog	pre!	BROAD1
<i>C. elegans</i>		WS 116 Apr 04
<i>C. briggsae</i>		cb25.aqp8 Jul 03

# Ensembl Species (2012 - partiel)



**Alpaca**  
*Vicugna pacos*  
vicPac1



**Anole lizard**  
*Anolis carolinensis*  
AnoCar2.0



**Armadillo**  
*Dasyurus novemcinctus*  
dasNov2



**Baboon** ([preview - assembly only](#))  
*Papio hamadryas*  
Pham



**Budgerigar** ([preview - assembly only](#))  
*Melopsittacus undulatus*  
MelUnd6.3



**Bushbaby**  
*Otolemur Gamettii*  
OtoGar3



**Ciona intestinalis**  
*Ciona intestinalis*  
KH



**Ciona savignyi**  
*Ciona savignyi*  
CSAV2.0



**Caenorhabditis elegans**  
*Caenorhabditis elegans*  
WBcel215



**Cat** ([preview new assembly Felis catus-6.2](#))  
*Felis catus*  
CAT



**Chicken** ([preview new assembly Galgal4](#))  
*Gallus gallus*  
WASHUC2



**Gibbon**  
*Nomascus leucogenys*  
Nleu1.0



**Gorilla**  
*Gorilla gorilla*  
gorGor3.1



**Guinea Pig**  
*Cavia porcellus*  
cavPor3



**Hedgehog**  
*Erinaceus europaeus*  
HEDGEHOG



**Horse**  
*Equus caballus*  
EquCab2



**Human**  
*Homo sapiens*  
GRCh37



**Hyrax**  
*Procavia capensis*  
proCap1



**Kangaroo rat**  
*Dipodomys ordii*  
dipOrd1



**Lamprey**  
*Petromyzon Marinus*  
Pmarinus\_7.0



**Lesser hedgehog tenrec**  
*Echinops telfairi*  
TENREC



**Macaque**  
*Macaca mulatta*  
MMUL 1



**Platyfish** ([preview - assembly only](#))  
*Xiphophorus maculatus*  
Xipmac4.4.2



**Platyplus**  
*Omithorhynchus anatinus*  
OANA5



**Rabbit**  
*Oryctolagus cuniculus*  
oryCun2



**Rat** ([preview new assembly Rnor 5.0](#))  
*Rattus norvegicus*  
RGSC3.4



**Saccharomyces cerevisiae**  
*Saccharomyces cerevisiae*  
EF4



**Sheep** ([preview - assembly only](#))  
*Ovis aries*  
oviAri1



**Shrew**  
*Sorex araneus*  
COMMON\_SHREW1



**Sloth**  
*Choloepus hoffmanni*  
choHof1



**Spotted Gar** ([preview - assembly only](#))  
*Lepisosteus oculatus*  
LepOcu1

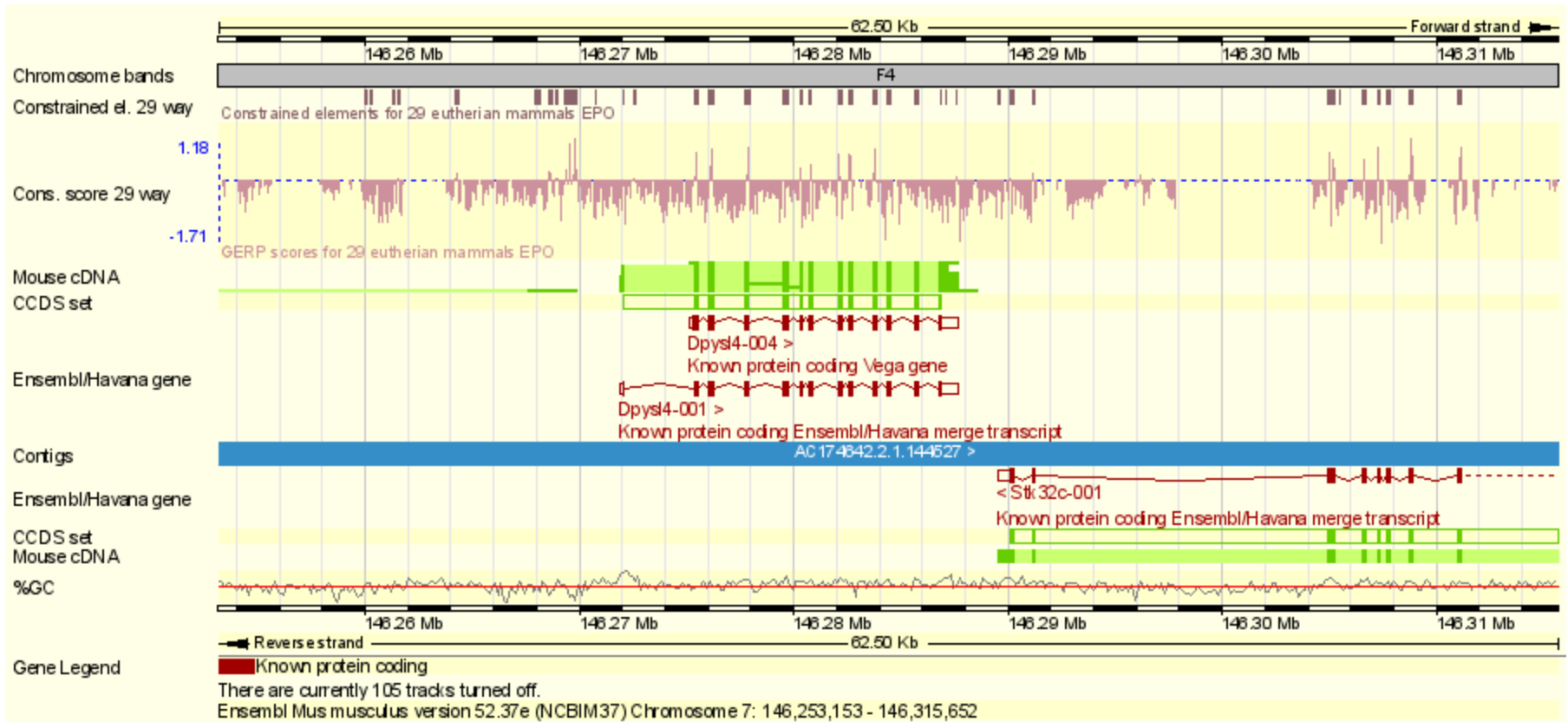


**Squirrel**  
*Ictidomys tridecemlineatus*  
spetri2



**Squirrel monkey** ([preview - assembly only](#))  
*Saimiri boliviensis*  
SaiBol1.0

# Ensembl: « contig view »



# Genome browser UCSC

– <http://www.genome.ucsc.edu/>

