



1809-2009



Deux siècles de science

Questions d'hier,
réponses d'aujourd'hui,
horizons de demain

Le séquençage du génome humain et après ?

Daniel Locker, Professeur émérite des Universités, CNRS, Orléans



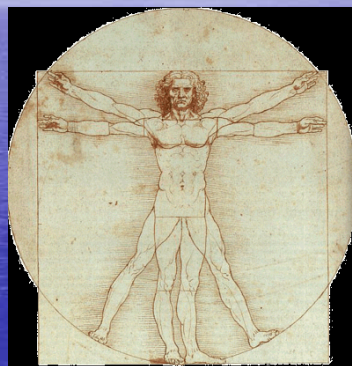
Comment a-t-il été séquencé?
Que nous apprend la séquence?
Et après?

D. Locker Muséum 13 mai 2009

2



Comment se présente le génome ?



D. Locker 30 septembre 2009

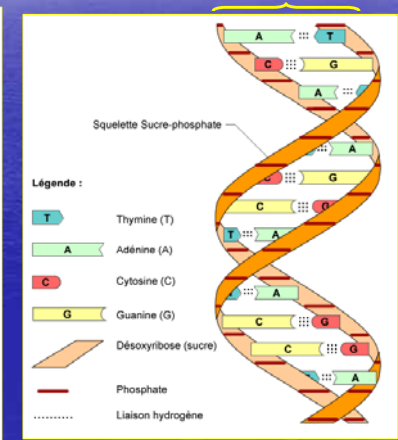
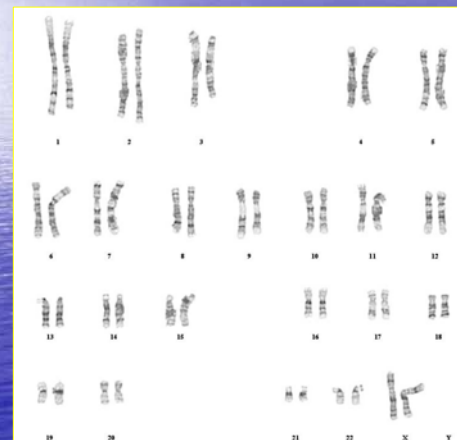
3



Le génome humain

* 23 paires de chromosomes
* 3 300 000 000 paires de bases

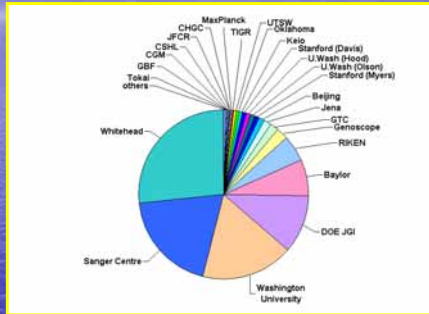
A :: T
1 paire de bases



D. Locker 30 septembre 2009

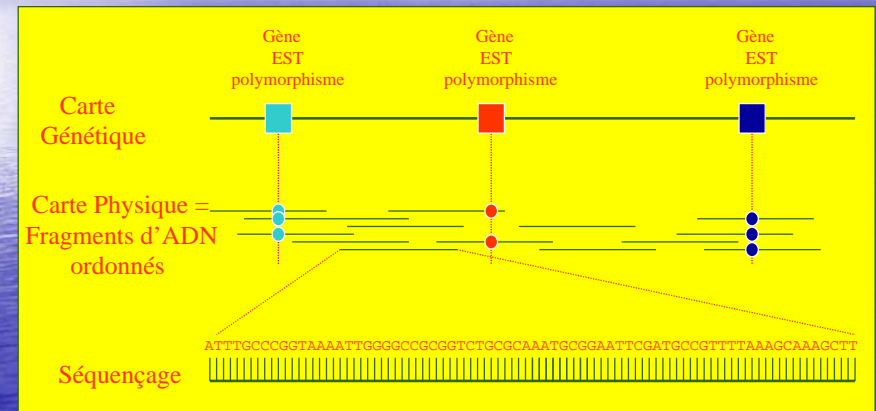
4

Les efforts de plusieurs laboratoires aboutissent au séquençage total du génome humain



D. Locker 30 septembre 2009

Les 3 étapes du projet de séquençage du génome humain par le consortium public:



D. Locker 30 septembre 2009

Le privé s'attaque aussi au séquençage du génome humain avec une nouvelle stratégie : le séquençage à la "mitrailleuse" !!!

D. Locker Muséum 13 mai 2009

Deux stratégies de séquençage

Stratégie de C. Venter Société Celera
«Shot-gun» = Séquençage aléatoire puis assemblage des séquences



C. Venter

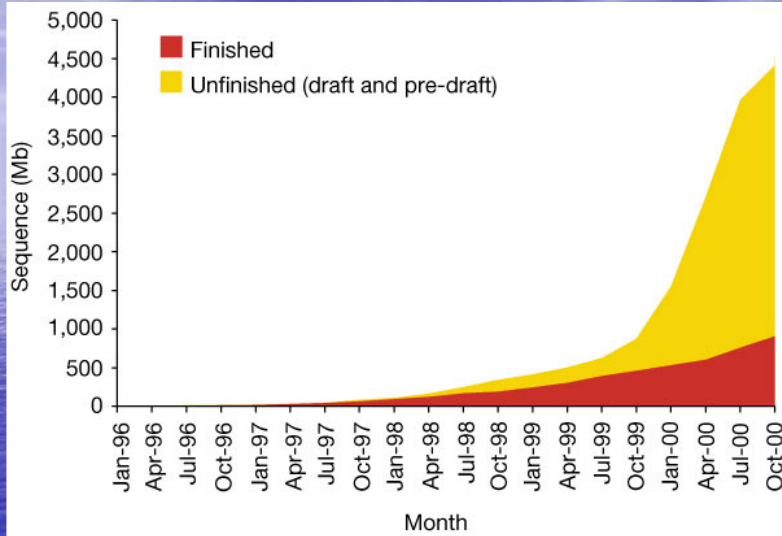
Stratégie du Consortium public
«Sequence-ready» = Production de cartes physiques prêtes à séquençer



F. Collins

D. Locker 30 septembre 2009

Les effets de la concurrence



D. Locker 30 septembre 2009

Février 2001 : premier brouillon de la séquence du génome humain



D. Locker 30 septembre 2009

```

ACGACCTCCATAGTTTGTATGTTGGCACGCCAGCGTTTATGTACATACCCGIGATCGTATGTTGTATGAAAGTAAACAGGCCAAATGTT
TATTAAGGCGCTTTTCCACACAC
TATAACGACTCTCTTCCGGCT
AATTTATCCGCCCGTTTGGAT
CGCGGAACCAACCTCATATA
TACCATACCACACCAAAA
TACAAACGACAAAACAATCC
AGGCCTAGACAAAAAACA
CGTGTCTCTTTTCTCTGT
GTACAAACACTCGGTAGAAC
ATTACATACATACGTACTTGT
GCCAATCCACGATGAAACAT
CGCGTTTTATTTAAACACA
ATGTATAGGTTGCCCGTTT
GGATATTTCCGCCACTGATGCTTTTCTCAGCCGAGTGGAAACITTTTGTGAATATATCTGCATACATACTACTACGTAGTACATGCAA
GGCCTGTACATCCATTTGGTCTCGCTTTTGTTTTGTAGTTAATAGTAAATCGGCATGTTACTAAATGCAAAAAGTGTACGCCCTATGTATGTCG
CAAGTATGCATACAAAACAAAAGCCACGCTGTACAATCAAAATGTAATTTTGGTGGCCATCCAGTCCGCCATTTTCTTATTGTTTGT
ATAAATATACAAATATGTACATTTATACCAAGAATCGTGAATGCCITGATGTAACCGTAAACCGCATTTGCTTTTCTTATGAAATATGC
TGGCAAAATCCGATCAAGTTTTTTTTCATCCGENOMEATTCGTAATAGATCAATGTCAAAAATAAGGAAAATATGTACATAAAAAACA
CATACATATAAAAACCTGTATGAAAAGCCAAATCACAGGCTTGGTAAATCCGAAATTTGGCTTATTTTATCCATGTTGCTATAAAA
TACAAATGAAGGCTGTGCAAAACCTTGGCTATCAAAATTTATGGCCACATGTATATTTCTGTGGTTACATACATATGTATCCCCAGTAAA
TGGCTCAATATTTATTTATTTATTTTCTTTTGGCAGTGTAGGAAAGTATTTTCTTTTCAATAATGGCATGCAAGATGGGGCTA
TTTTCCGCAAAAAGAAAGTTCCTCAGTTTGTAACTGTGGCCCTTGGTTTTCACCTTTAAGCCCTTGGAAATCGAATCGAAATGAAGACA
ATCCAAATCCGATTCACAAAACCGGAACCGGTCTACTGCTCTGCTTCTTCTTCTCACAACATAGTTCCACTCTTGGCCGTTCAATGTA
AAAGCTTTTTTATCTATGGTTAACAGCTGGTGCCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT
TTCAATCGGAAACAGAGAATTCACCTTTGGCAGGATTTAGTATTAATTTGTTGGTATTTGACATCAATTTATCGGTATCTGTAGATAACGA
GTATCTACATATTTTATTTACATGGAAAGTAAATAAAAAGGATAACGCCGCCAGTTCACACAGATCTTTTTTTTGAAGTATTAATTTAGG
GAGATTTTGAAGTCTATTTAAGCTTTTTTTTTTTAGGCAACAATATTTTTTCACTTAATAAATAAATTTAAAGAGCTTAGCTTAGCATACT
AAAATTCAGTTTTTACAGTGTAGAAGTGTAAATTTTAAACAATCAGCAATGTGTAACGCAATGTAGAGAAAAATCGCTGCATTCGGTAT
TTATTTAATTTAGTCAATTTTGTGAGCAAGTGTCAACAAAACATCAGGCACTCAATTAACCCGCAATACTCAAAAACAAAATGACAGTTTT
TCACTTTGTTAGTATTAAGSACATCTTAGTGGTTAGTGGATGCCAGACATGAGGTCAGCCAAATGATTTAGTGTATTTGTTAGCCCC
ATCGACTGGCATCCGCCACTAATAATGCTAATGAACATCAACCAACTATGATCAGCAAAACAATCAACACTGGGTTAATTTGCCCTTGG

```

Le côté obscur du génome

D. Locker 30 septembre 2009

Que faire de cette masse d'information ?

Il faut identifier tous les gènes et leurs fonctions : c'est l'étape d'annotation de la séquence



D. Locker 30 septembre 2009

```

ATCCCTCCATAGTTTGTATGTTTGGCACGCCAGCGTTTATGTACATACCCGIGATCGTATGTTTGTATGAAAGTAAACAGGCCAAAT
TATTAAGGCGTTTTCGCCACACCGTGTATGTACATATGATGTATATGACACTGAGCGCTCTCTGACGACACACTAATTTCTACCAACAATAT
TATAACGACTCTCTTCGCGCTCTTCGCTCTCCGCCACACACGACACCCAGCAAGAATTTGGCACACAGTCCAACGCCAACGCCAAGCAAT
AATTTATCCGCCGTTTGGATTTGGGAAGCACAGTGAAGGTAAGGCAGAGCGCAATTCATTAATTCATTCGTTGTGGTACGAGCGG
CGCGGAACCAACCTCATATATGTTATTTTCGCAATTCGCAATGATACGATACCCATACCGTGTGCTGTACIGTGTAACGTATAGCATATA
TACCATACCAACCAACAAACAAGCGATACGATACGATACAAATAAATCAGCGTATGCGAGCGCTGTTTGTATTTTTTTTCAAGCTAAA
TACAAACGCAAAACAATCCGCAAGCCGCGCGGTTTTCGCGTTTTCACCAACCAAAACAATAAATCAAAACAAACAATAAGCGAAAA
AGCCCTAGACAAAAAAGAAAGTACAAAAACAACGTCATGTGAAATGTTAAATTTTCGAATTCATTCGAAGTTTGGCTGTGTC
CGTGTCTTTTTTTTCCCTGTAAAGTGCATCATATTCGCAATTTGTGTCGCTACATGCATACATACATCAATCAATCAATACAAATAGCAG
GTACAAACAATGCGTAGAACAGTATTAATAAAGGAAAAACGCAAAAAAATCGGCTTAAGACTCGACTCCAATTTGCGTGTGTGTGCGT
ATTACATACATACGTAATCTGTGCTAAACTAAATATATATGCTTTAACCATTTAGGAATAAAGTGAAGATATAGAACGGAAATCCCGATA
GCCAATCCAGATTTGAAACATTCGGACTAGATCCGAAATCTCACAGTTTATGCAACACTTTTCATCAATACAGTAAAGCTGTGTGTCAGGTATA
CGCGCTTTTATTAACAACAACGCCACGACACACAACAACAATTTATATACACTGATGGCAACAACAACAATCATGSGCGGTTT
ATGTATAGGTTGCGCGGTTTAAATCATTTTTCCGCCAAACGCAATGTATAATTAACAATAAAGCGAGGCACTTATTTATTTATTTATTT
GGATATTTCCGCCACTGATTTGCTTTCTCCGCGCGAGTGGAAACTTTTGTGAATATATATCTGATACACTACTACTAGTACATGCAA
GCGCTGTACATCCATTTGGCTCTGCTTTTGTTTTTAGTTATTAAGTAAATCGGCATGTTACTAAATGCAAAAAGTACGCCCTATGTATGTG
CAAGTATGCATACAAACAATAAAGCGCGCTGTACAATCAAAATGTTATTTTGGCGGCAATCCAAGTCCGCCATTTTTCTTATTTAGTTTGT
ATAAATATACAAATATGTACATTTATACCAAGAAATCGTGAATGCCTTGATGTAAACGGTAAACCGCATTTGCTTTTGTGTTATGAAATATG
TGGCAAAATCGCATCAGTTTTTTTTGCAATGCTGENOMEATTACGTAATAGATCAATGTGCAAAAATAAGGAAATATGTACATAAAAAAAA
CATACATATAAAACCTGTATATGAAAGACCCAAATCACAGCTTGCATATCCGAAATTTGCTTATTTTTTATCCATGTGTGCATAAAA
TACAAATGAAGCTGTGCAAACTTGGCTATCAAAATTTATGCGCACATGTATATTTCTGTGTTACATACATATGTATCCCCAGCTAAA
TGGCTTAATATTTATTTATTTTTTTTTTTTGGCAGTTGTAGGAAAGTATTTTTGCTTTTTTCAATAATGGCATGCAAGATGGGCTA
TTTTGGCAAAAAGAAATGGTTCTTCATGTTTTGTAACTGTGGCGCCCTGGTTTTGCACTTTAAGCCCTGGAATGGAAATCGAAATGAAAGCA
ATCCAAATCCGATTCACAAAACCGGAACCGGCTACTGCTGCTGCTTCTTCTTCTCCAAACATAGTTCCACTCTCTGCGCGTTTCATGTAAA
AAAGCTTTTTATCTATGGGTTAACAGCTGGTGTATTAATACATTACTTTGTAACAACAACAGCTAGATTAITGACATTGAAATCCCGG
TTCAATCGGAAACAAGGAATTCATTTGGACAGGATTTAGTATTAATTTGTGTTGTTGCACTAATTTATCGGATCTGTAGATAACGA
GTATCTACATATTTATTTACATGTAAGTAAATAAAGGATAAGCCCGCAGTTCCACACAGTCTTTTTTTTTAGAGTATTAATATGAA
GAGTTTTAGAAATGCTATTTAAGCTTTTTTTTTAGGCAACATATTTTTACACTTAATAAATAATTTAAAGAAGCTTAGCTTAGCATACT
AAAATTCAGTTTTACAGCTGAGAATGTTAAATATTTAAACAATCAGCAATGTGTAACGCAATGTAGAGAAAAATCGCTGCATTCGGTAT
TTATTTATTTAGTCAATTTGTGAGCAAGTAGTCAACAAAAACATCAGGCATCTAATTAACCGCAATCTCACAAACAATAAATGCAAGTTT
TCATTTTGTGTTAGTATTAAGGACATCTTAGTGGTTAGTGGATGCCAGACATCGAGGTCAGCCAAATGATGTTAGTATTTGTTGAGCC
ATCGACTGGCATCCGCCACTAATAATGCTAATGAACATCAACCAACTATGATCAGCAACAATTAACACTGGGTTAATTTGCGCTGGG

```


Les surprises du dénombrement des gènes

Lors de l'annotation du génome humain, on trouve seulement 23-28 000 gènes au lieu des 100 000 attendus

C. elegans 18 200 gènes

D. melanogaster 14 000 gènes

H. sapiens 23-28 000 gènes.


D. Locker 30 septembre 2009

Les conclusions (provisoires) de l'annotation du génome

- * Seulement 1% du génome code des protéines
- * Il y a de 23 à 28 000 gènes dans le génome humain
- * Les exons représentent ~5% de chacun des gènes
- * Les gènes représentent ~25% du génome humain
- * ~60% des gènes présentent une possibilité d'épissage alternatif ; ils peuvent coder ~60 000 protéines différentes.
- * Les séquences répétées constituent plus de 50% de notre patrimoine génétique. Ce sont surtout des transposons non fonctionnels.
- * Dans les chromosomes, de grandes régions sont dupliquées plusieurs fois.

D. Locker 30 septembre 2009

Quelles sont les retombées du séquençage du génome humain?

Des retombées positives:

- 1) Dans le domaine médical, découverte de nouveaux gènes impliqués dans des maladies
- 2) Études des caractères quantitatifs
- 3) Mise en place de la médecine prédictive
- 4) **Apport d'informations importantes sur l'évolution**

D. Locker 30 septembre 2009

Les retombées négatives des résultats du séquençage du génome humain:

- 1) Le typage des individus
- 2) Le brevetage des gènes

A la recherche des gènes impliqués dans les maladies humaines en utilisant le polymorphisme de l'ADN : exemple des SNP

```

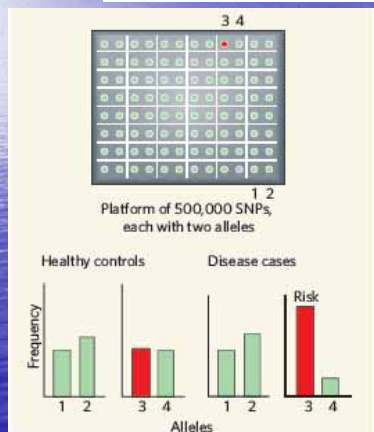
ATGGTAAGCCTGAGCTGACTTAGCGT-AT
ATGGTAAACCTGAGTTGACTTAGCGTCAT
          ↑           ↑           ↑
          SNP       SNP       indel
    
```



Utilisation des SNP dans le cadre du projet GWA

GWA (Genome Wide Association) ou la ruée vers l'or

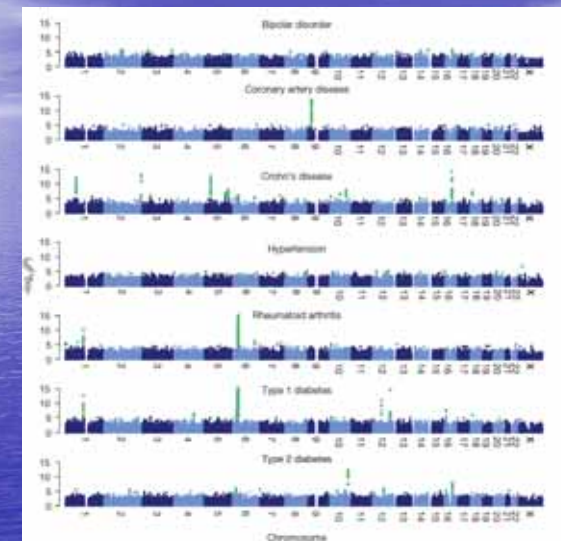
In a tour-de-force demonstration of feasibility, a consortium of 50 research teams uses 500,000 genetic markers from each of 17,000 individuals to identify 24 genetic risk factors for 7 common human diseases.



Etude de la distribution de 500 000 SNP dans 17 000 cas correspondant à 7 maladies, comparaison avec 3000 individus témoins

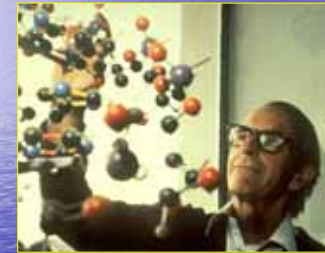
Vol 447 | 7 June 2007 | doi:10.1038/nature05911

Les résultats du consortium GWA

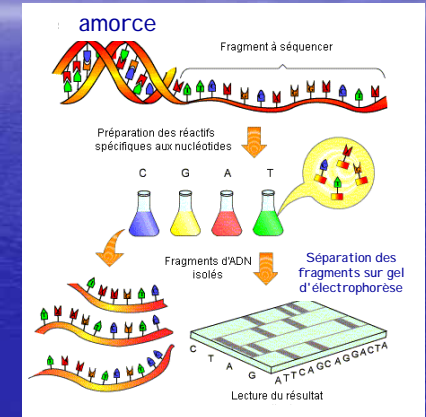


Des améliorations techniques importantes vont permettre le séquençage des génomes individuels pour un faible coût. On pourra ainsi prédire certaines maladies

La technique initiale : le séquençage par la méthode de Sanger



F. Sanger deux fois prix Nobel de chimie 1950 et 1980

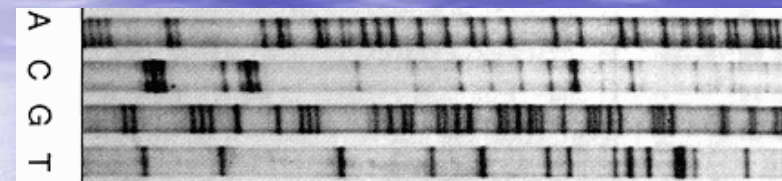


Première amélioration technique de la méthode de Sanger vers 1990 : l'électrophorèse capillaire

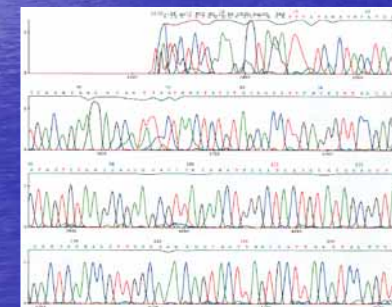
- Automatisation de la méthode :
 - Ajout sur les nucléotides de fluorochromes de différentes couleurs.
 - Lecture par un laser des fragments qui migrent sur le gel.
 - Stockage des données dans la mémoire de l'ordinateur.



Lecture des séquences d'ADN avant l'électrophorèse capillaire



Lecture des séquences d'ADN avec l'électrophorèse capillaire

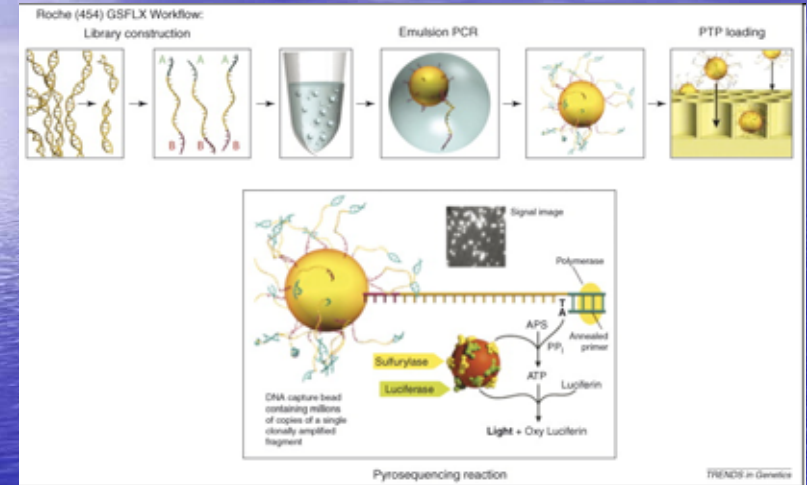


Une nouvelle méthode : le séquençage en masse par pyroséquençage

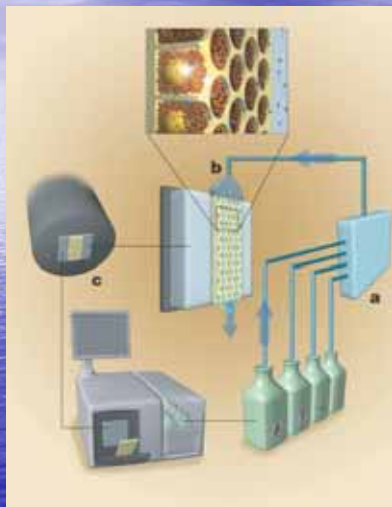
Le pyroséquençage ou que la lumière soit!



Principe du pyroséquençage



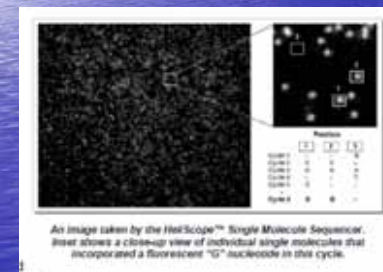
Pyroséquençage en masse



Une caméra reliée à un ordinateur lit simultanément la lumière émise par plusieurs millions de billes. Ainsi un génome bactérien est séquençé en 7h.

2009 les dernières nouveautés: le séquençage des molécules uniques d'ADN

Principe:
L'ADN polymérase et un nucléotide marqué par fluorescence (C, G, A ou T) sont ajoutés aux molécules uniques d'ADN.
Lecture de la fluorescence
Coupeure du groupement fluorescent
Le processus se poursuit avec les trois autres bases.



Diminution très rapide du coût du séquençage des génomes



2001 premier génome humain séquencé par la technique de Sanger coût 500 millions de \$(Personnel : consortium de laboratoires)



2007 la société 454 Life Sciences séquence le génome de J.D. Watson par la technique du pyroséquençage coût 1 million de \$(Personnel : 200 personnes)



En 2009 la société Heliscope séquence le génome de S. Quake par la technique du séquençage de molécules uniques d'ADN, coût 50 000 \$(2 personnes)



Prédiction pour 2013 un coût de 1000 \$ pour le séquençage d'un génome individuel avec 1 appareil automatisé



29



Les apports du séquençage des génomes individuels

30



Médecine prédictive

The diploid genome sequence of an Asian individual

Table 3 | Number of alleles identified that increase the risk to specific complex diseases

Traits	Associated genes	Associated SNPs	Predisposing alleles in YH	
			Number	Per cent
Alzheimer's	7	16	9	56.3
Diabetes	26	46	7	15.2
Hypertension	8	10	1	10.0
Obesity	6	27	1	3.7
Parkinson's	7	11	1	9.1
Hypolactasia	1	2	0	0.0
Alcohol addiction	3	3	0	0.0
Tobacco addiction	7	19	12	63.2

Nature Vol 456 | 6 November 2008

31



Les gènes impliqués dans le cancer, exemple de la leucémie avec la découverte de 8 nouveaux gènes impliqués dans la maladie

DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome

Table 2 | Non-synonymous somatic mutations detected in the AML sample

Gene	Consequence	Type	Selexia tumour reads WT-variant	Selexia skin reads WT-variant	Conservation score of mutant base	Mutations in other AML cases*
CDH24	Y590K	Nonsense	9/9	16/0	0.998	0/187
SLC15A1	W77X	Nonsense	15/12	19/0	1.000	0/187
KND1C1	L799F	Misense	7/8	20/0	NA	0/187
PTPRF	P1235L	Misense	9/13	16/0	1.000	0/187
GRIN1B	R176H	Misense	15/10	14/0	NA	0/187
GPR123	T38I	Misense	11/11	13/0	NA	0/187
EBI2	A338V	Misense	7/12	18/2	1.000	0/187
PCDC	P1004L	Misense	19/9	15/1	0.98	0/187
FLT3	ITD	Indel	18/12	8/0	NA	51/185
NPM1	CATG ins	Indel	36/6	33/0	NA	43/180

Nature Vol 456 | 6 November 2008

32

Etude des origines géographiques des populations humaines

Geographic map and Venn diagram of five sequenced genomes.



Nature Vol 460|20 August 2009

33

D. Locker 30 septembre 2009

La pharmacogénétique

Single-molecule sequencing of an individual human genome

Après séquençage du génome d'un chercheur :

Mauvaise nouvelle, risque de maladie cardiaque augmenté

Bonne nouvelle, prédisposition à bien réagir aux statines, une classe de médicaments destinés à faire baisser le taux de cholestérol.

Le séquençage de son génome indique aussi qu'il est porteur d'un gène impliqué dans la dégradation du caractère au fil des ans. Commentaire du chercheur «*Ce n'est pas une révélation ma femme aurait pu vous le dire*».



Nature Biotechnology 10 august 2009

34

D. Locker 30 septembre 2009

A la maternité dans un futur proche

Birth in the age of genetic mapping...



35

D. Locker 30 septembre 2009

Merci de votre attention



D. Locker 30 septembre 2009

36