Colloque EPGV_2009
11-12-13 mai

Les outils de séquençage / re-séquençage
Introduction

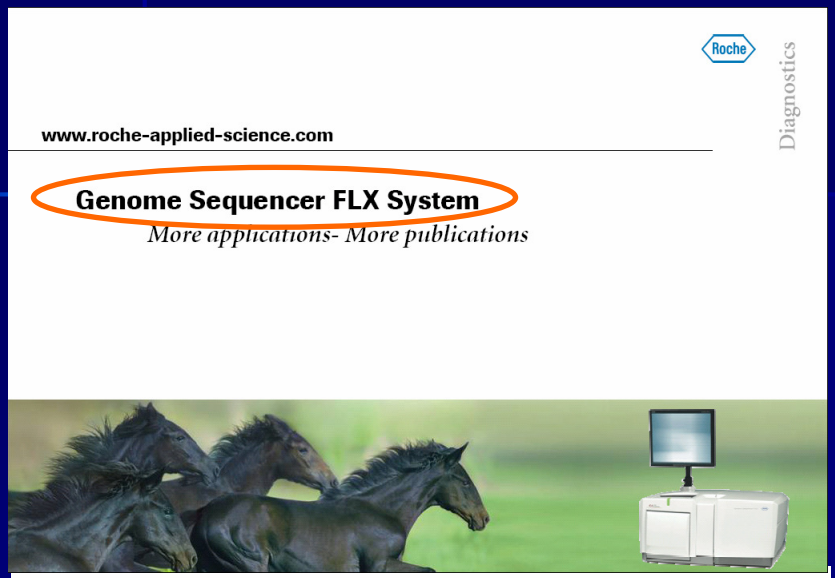
Marie-Christine Le Paslier

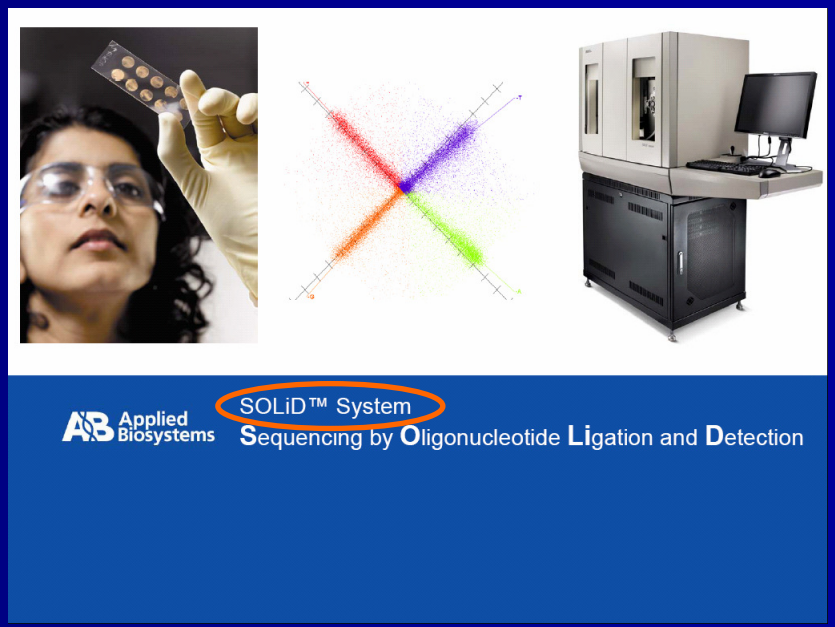
E P G V
Etude du Polymorphisme des Génomes Végétaux
UR1279_DGAP

NTS : Nouvelles Technologies de Séquençage
NGS : Next Generation Sequencing

=>1000 $/génome humain

« 2d generation technologies »
« massively parallel sequencing »



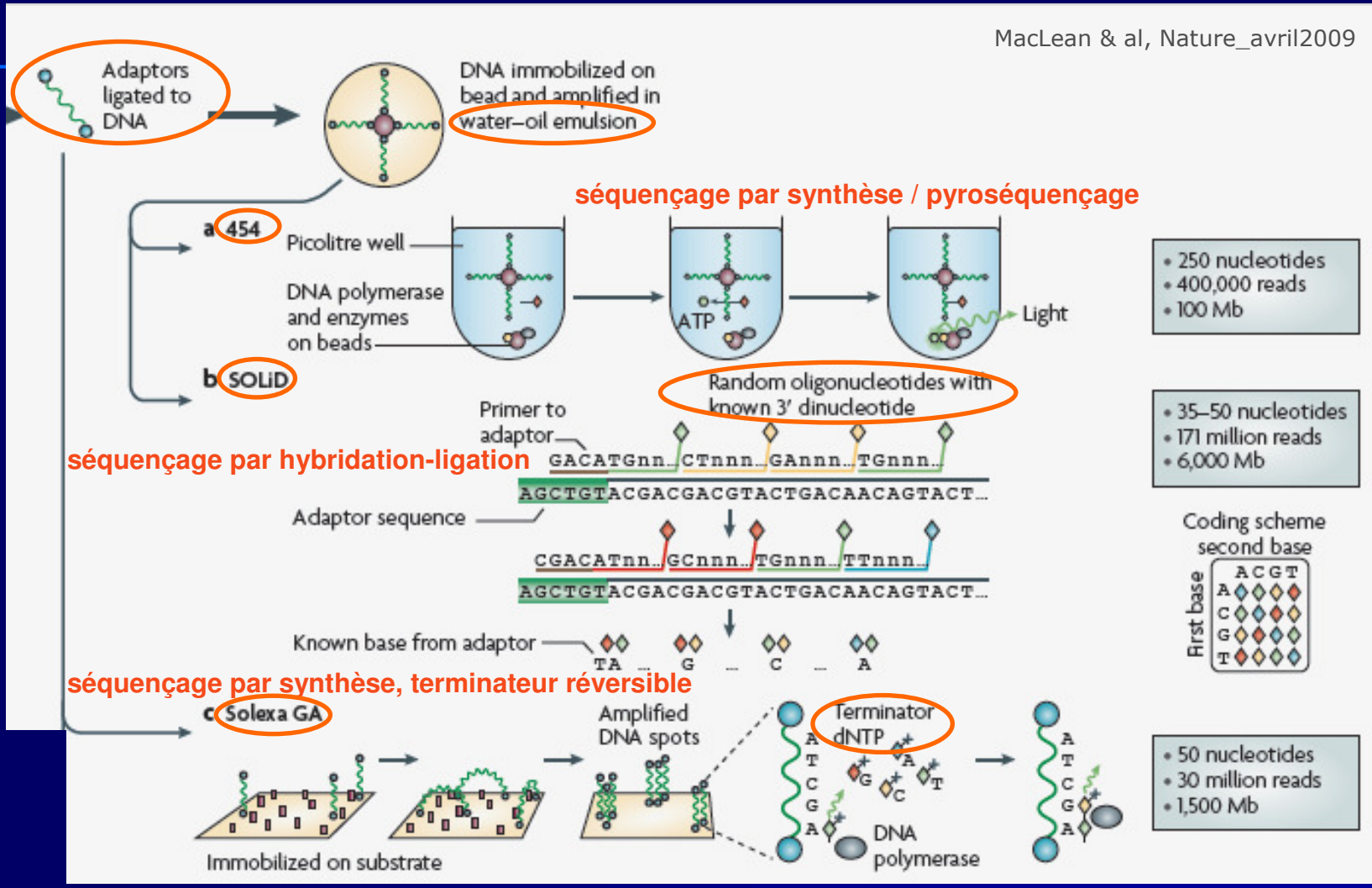


Les 3 technologies implantées

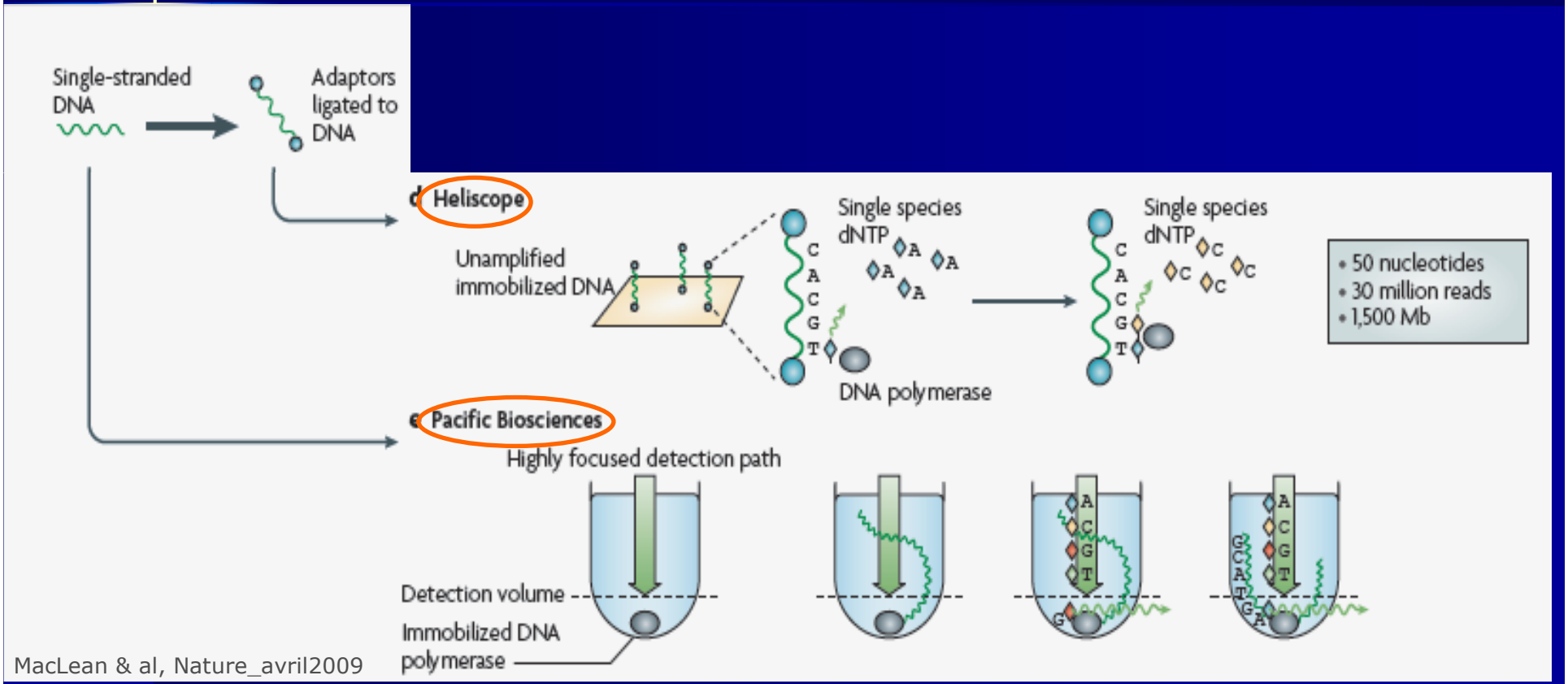
GS-FLX (Roche), GA (Illumina), SOLiD (ABI)

- Préparation de la banque de molécules
  - fragmentation (nébulisation, ultrasons)
  - ligation d'adaptateurs aux extrémités
  - isolement du simple brin
- Amplification clonale de chaque molécule (PCR)
- Séquençage par synthèse ou hybridation-ligation

Les 3 technologies implantées
GS-FLX (Roche), GA (Illumina), SoLid (ABI)



Les technologies en cours d'implantation



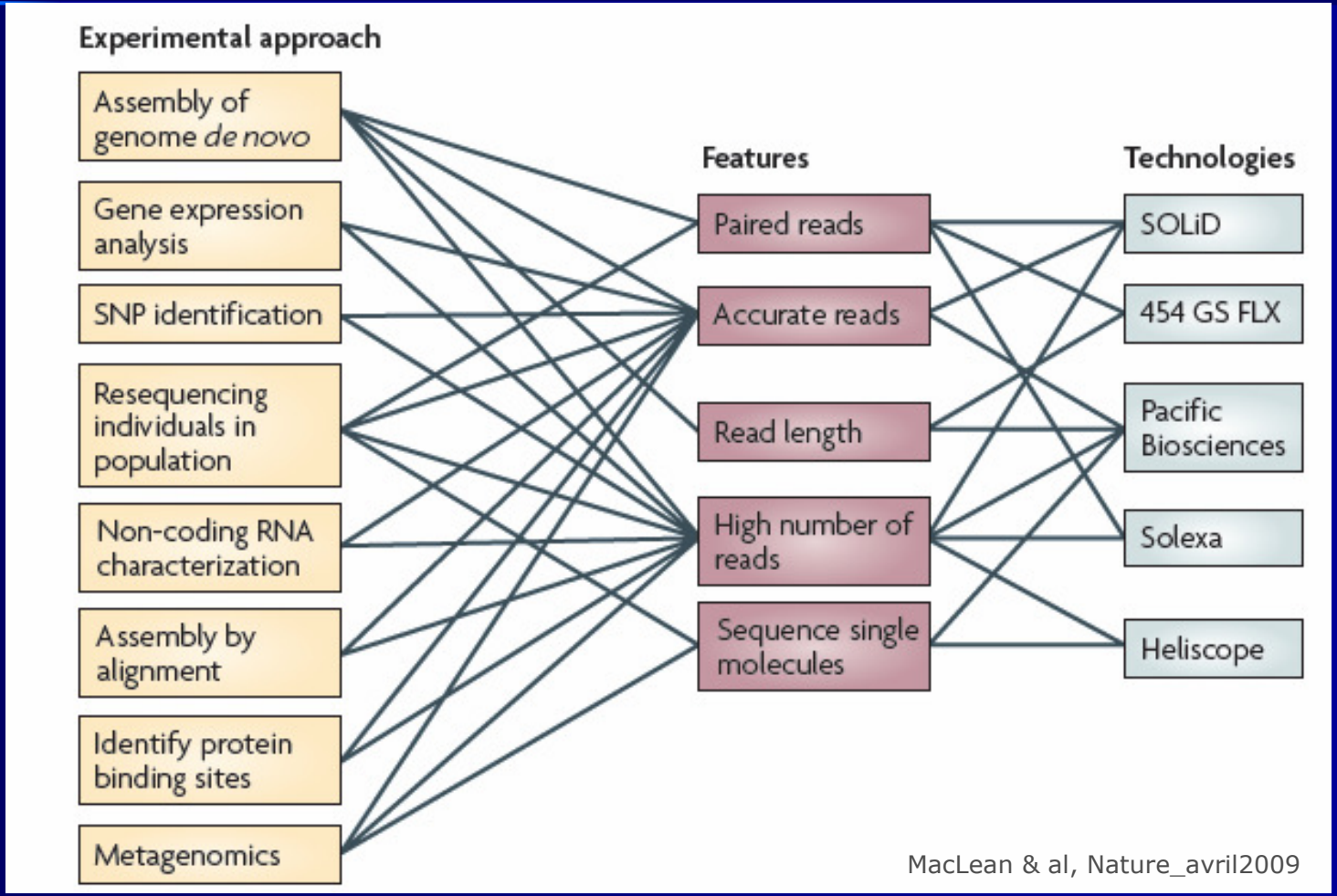
MacLean & al, Nature_avril2009

et … à venir ! … nanopores …

Comparatif GS-FLX / GA

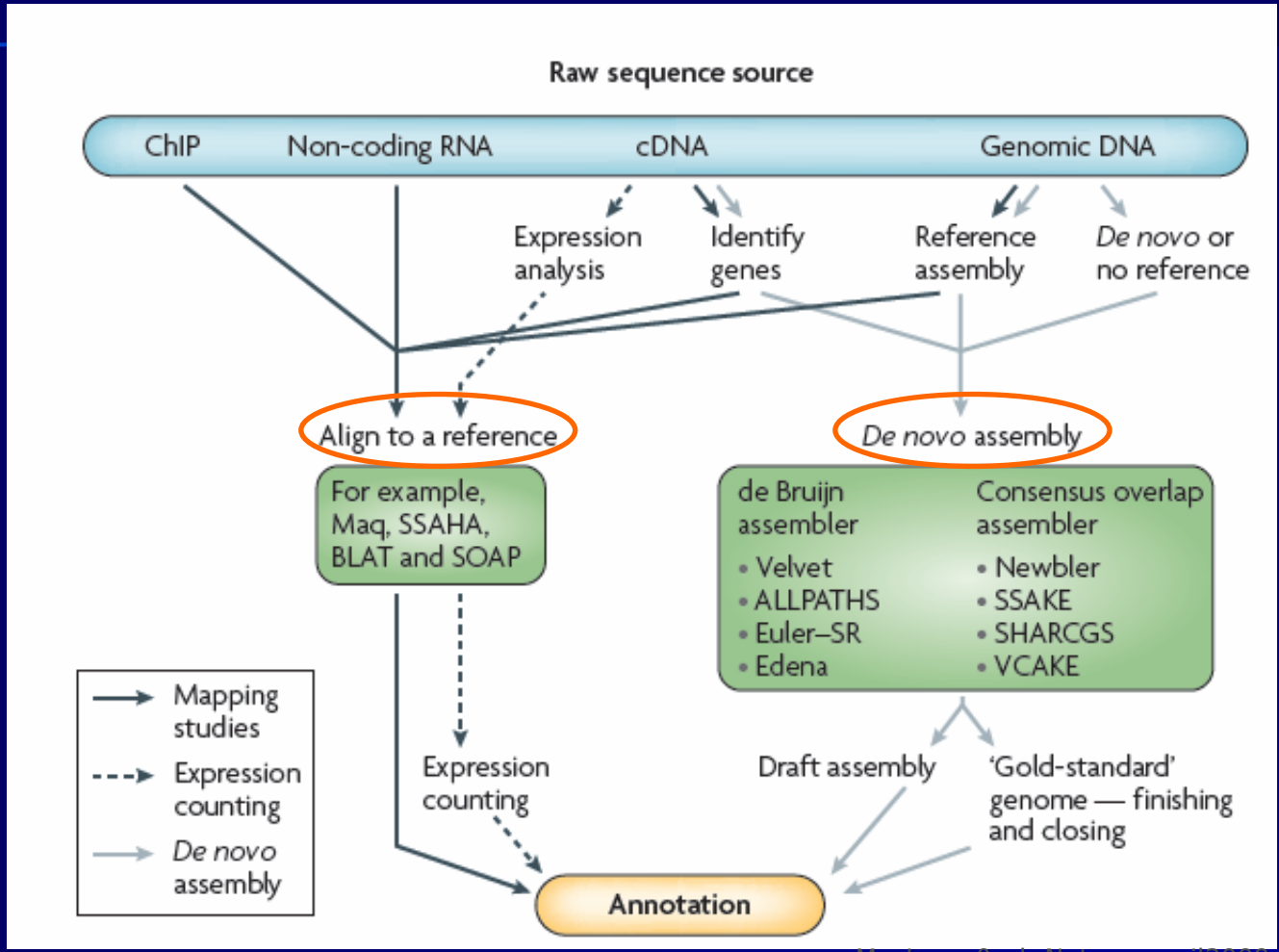
	GS-FLX 454 Roche		GA Solexa Illumina	
	Standard	Titanium	GA II	GA II
Lectures/run	400 x 10^3	1 x 10^6	80-100 x 10^6	80-100 x 10^6
Longueur de lecture	250 bases	400 bases	35 bases	75/100 bases
Bases/run	100 Mb	400-600 Mb	3 Gb	7,5/10 Gb
Tps de run (simple lecture)	7,5 h	10 h	3-5 jrs	4-9,5 jrs /
Flexibilité	2-4-8-16 fractionnements (avec réduction nb lectures liée au masque)		8 canaux	
Extrémités appariées Lectures appariées	3kb, 8 kb, 20 kb		200 à 5 kb	
Multiplexage /adressage	12 tags		12 tags	
Avantages / Inconvénients	Longueur / Homopolymères		Profondeur / assemblage	
Evolutions	lecture à 500→1000 bases automatisation des étapes de préparation		Lecture à 125 puis 150 bases	

Que peut-on réaliser avec ces technologies ?
Quelles technologies pour quels besoins ?



MacLean & al, Nature_avril2009

Les outils bio-informatiques



MacLean & al, Nature_avril2009

Notre utilisation des nouveaux outils

- Exploration des NTS initiée fin 2008 : PlantReSeq, AIP Bio ressources
  - Mutualisation de 7 équipes : exploration différente selon la question posée pour les différentes espèces végétales / matériels différents :

    ADN génomique (représentation) ou LR-PCR, ARNs
  - 2 outils : GS-FLX (454, Roche) et GA (Solexa, Illumina)
    - premiers résultats sur le blé : construction d'un index de séquences répétées du Chr 3B
    - en cours, l'analyse différentielle des séquences de pool d'amplicons de 10 kb pour 2 accessions de tomate
- Développements : réduction génomique par « Sequence Capture »
  - Nimblegen (en cours pour la vigne) : hybridation sur puce
  - Agilent : hybridation liquide
  - MIP (molecular inverted probes)

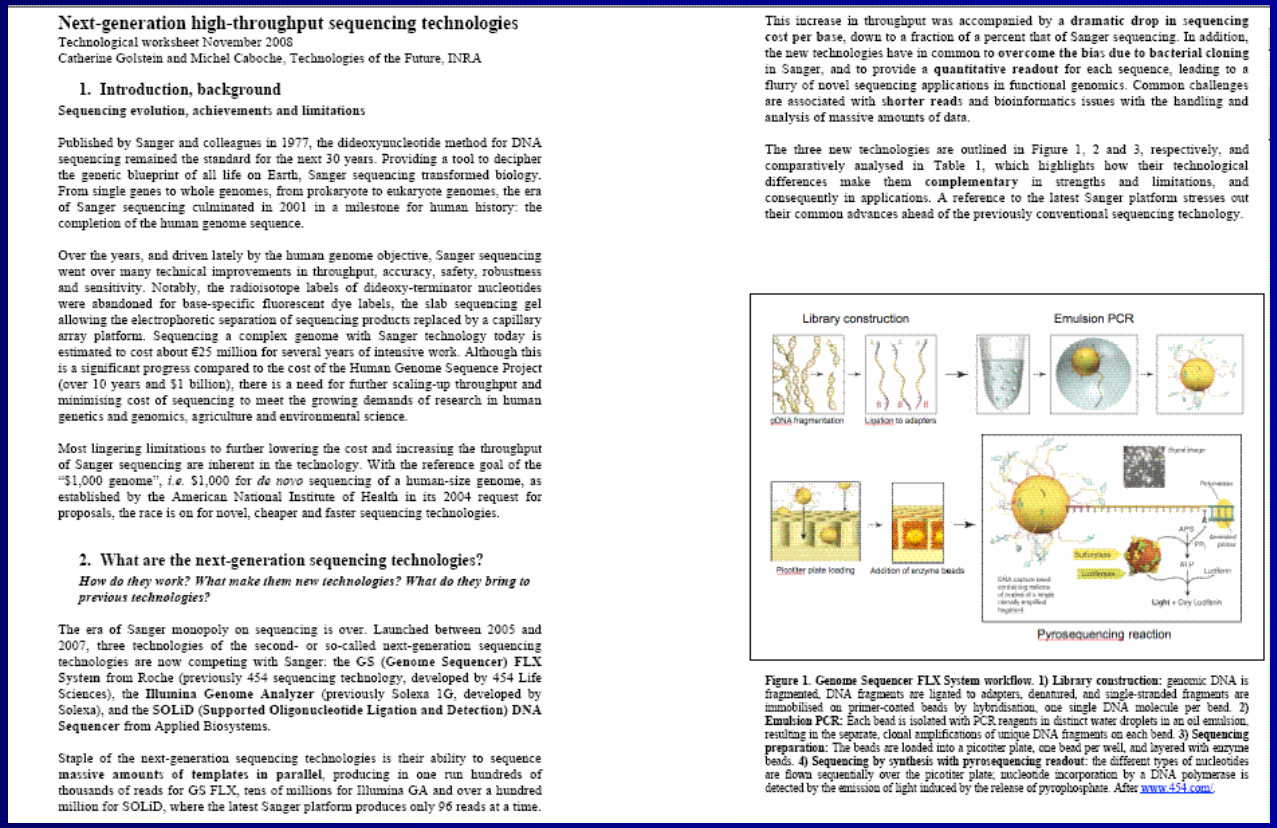
Quelles nouvelles questions scientifiques vont pouvoir être traitées avec ces outils ?

Document de l'unité Prospective de l'INRA :

« Next-generation high-throughput sequencing technologies »

Technological worksheet, Novembre2008

Catherine Golstein et Michel Caboche, Technology of the Futures, INRA



Next-generation high-throughput sequencing technologies

Technological worksheet November 2008

Catherine Golstein and Michel Caboche, Technologies of the Future, INRA

1. Introduction, background

Sequencing evolution, achievements and limitations

Published by Sanger and colleagues in 1977, the dideoxynucleotide method for DNA sequencing remained the standard for the next 30 years. Providing a tool to decipher the genetic blueprint of all life on Earth, Sanger sequencing transformed biology. From single genes to whole genomes, from prokaryote to eukaryote genomes, the era of Sanger sequencing culminated in 2001 in a milestone for human history: the completion of the human genome sequence.

Over the years, and driven lately by the human genome objective, Sanger sequencing went over many technical improvements in throughput, accuracy, safety, robustness and sensitivity. Notably, the radioisotope labels of dideoxy-terminator nucleotides were abandoned for base-specific fluorescent dye labels, the slab sequencing gel allowing the electrophoretic separation of sequencing products replaced by a capillary array platform. Sequencing a complex genome with Sanger technology today is estimated to cost about €25 million for several years of intensive work. Although this is a significant progress compared to the cost of the Human Genome Sequence Project (over 10 years and $1 billion), there is a need for further scaling-up throughput and minimising cost of sequencing to meet the growing demands of research in human genetics and genomics, agriculture and environmental science.

Most lingering limitations to further lowering the cost and increasing the throughput of Sanger sequencing are inherent in the technology. With the reference goal of the "$1,000 genome", *i.e.* $1,000 for *de novo* sequencing of a human-size genome, as established by the American National Institute of Health in its 2004 request for proposals, the race is on for novel, cheaper and faster sequencing technologies.

2. What are the next-generation sequencing technologies?

How do they work? What make them new technologies? What do they bring to previous technologies?

The era of Sanger monopoly on sequencing is over. Launched between 2005 and 2007, three technologies of the second- or so-called next-generation sequencing technologies are now competing with Sanger: the **GS (Genome Sequencer) FLX System** from Roche (previously 454 sequencing technology, developed by 454 Life Sciences), the **Illumina Genome Analyzer** (previously Solexa 1G, developed by Solexa), and the **SOLiD (Supported Oligonucleotide Ligation and Detection) DNA Sequencer** from Applied Biosystems.

Staple of the next-generation sequencing technologies is their ability to sequence **massive amounts of templates in parallel**, producing in one run hundreds of thousands of reads for GS FLX, tens of millions for Illumina GA and over a hundred million for SOLiD, where the latest Sanger platform produces only 96 reads at a time.

This increase in throughput was accompanied by a **dramatic drop in sequencing cost per base**, down to a fraction of a percent that of Sanger sequencing. In addition, the new technologies have in common to **overcome the bias due to bacterial cloning** in Sanger, and to provide a **quantitative readout** for each sequence, leading to a flurry of novel sequencing applications in functional genomics. Common challenges are associated with **shorter reads** and bioinformatics issues with the handling and analysis of massive amounts of data.

The three new technologies are outlined in Figure 1, 2 and 3, respectively, and comparatively analysed in Table 1, which highlights how their technological differences make them **complementary** in strengths and limitations, and consequently in applications. A reference to the latest Sanger platform stresses out their common advances ahead of the previously conventional sequencing technology.

Figure 1. Genome Sequencer FLX System workflow. 1) Library construction: genomic DNA is fragmented, DNA fragments are ligated to adapters, denatured, and single-stranded fragments are immobilised on primer-coated beads by hybridisation, one single DNA molecule per bead. **2) Emulsion PCR:** Each bead is isolated with PCR reagents in distinct water droplets in an oil emulsion, resulting in the separate, clonal amplifications of unique DNA fragments on each bead. **3) Sequencing preparation:** The beads are loaded into a picotiter plate, one bead per well, and layered with enzyme beads. **4) Sequencing by synthesis with pyrosequencing readout:** the different types of nucleotides are flown sequentially over the picotiter plate, nucleotide incorporation by a DNA polymerase is detected by the emission of light induced by the release of pyrophosphate. After www.454.com/.

sera mis en ligne sur http://colloque.inra.fr/epgv