FASTERIS

The Impressive Increase in Throughput of the illumina Genome Analyzer, as Seem from an User Perspective

Laurent FARINELLI

31 August 2009

Illumina Seminar

55° Congresso Brasileiro de Genética

Águas de Lindóia, Brasil

Revolution in Throughput

One run, one sample ..

Capillary Sequencer

1 gene

1 page

3'000'000x

Village

1'000 inhabitants

Illumina Genome Analyzer

A genome

A library with 12'000 books of 500 pages each

Earth

7'000'000'000 inhabitants

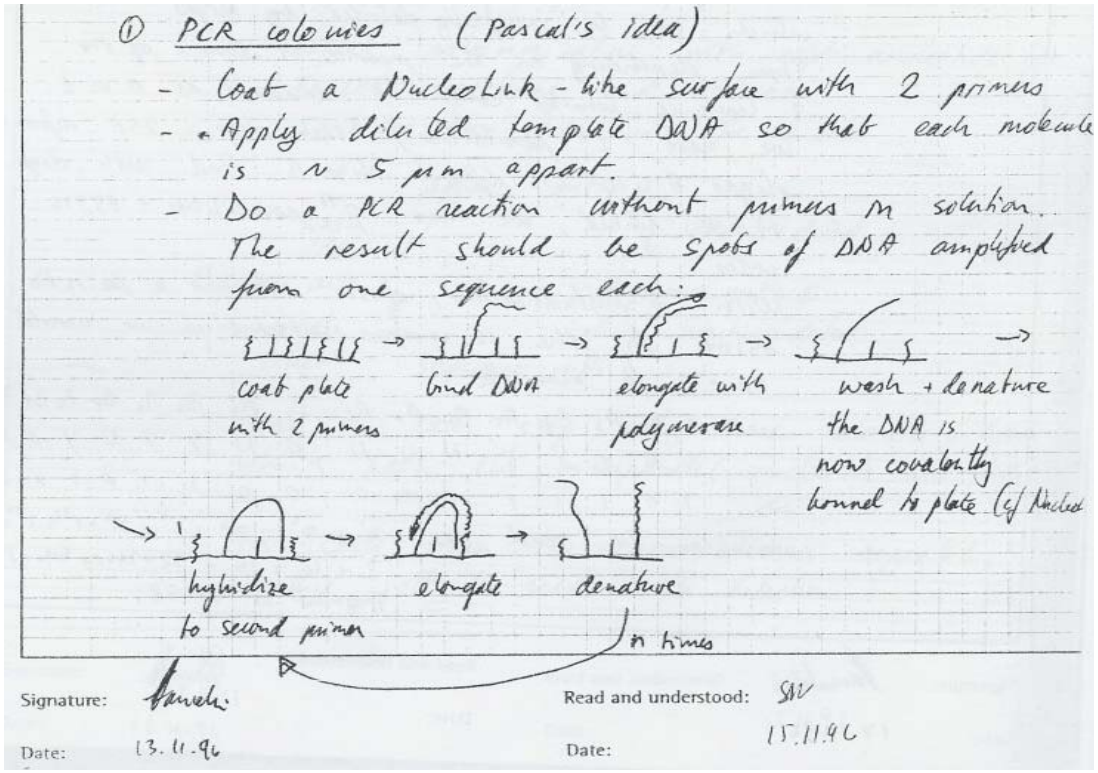
1996, early days at Glaxo..

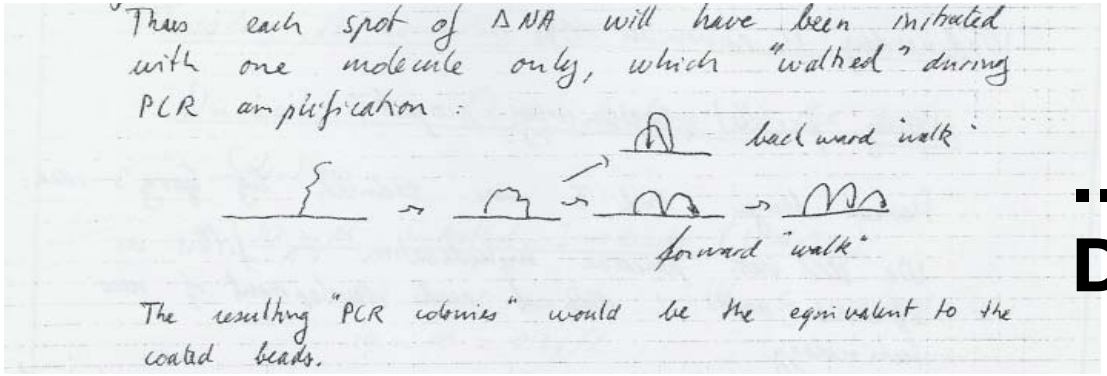
- Whole genome sequencing of *Streptococcus pneumoniae*
- *7 instruments ABI 377 glass plates*
- *700 sequences per day, reads of 500 bases*

⇒ *4 bases per second*
combined throughput

=> almost 1 year to sequence 2.1 Mb

1996, we need something else..





DNA Colonies..

1996-1997: **GlaxoWellcome**'s Geneva Biomedical Research Institute

Mayer P., Farinelli L. and Kawashima, E., 1997, Patent application WO 98/44151

.. now know as DNA Clusters

...Massively parallel sequencing

2003: Foundation of Fasteris

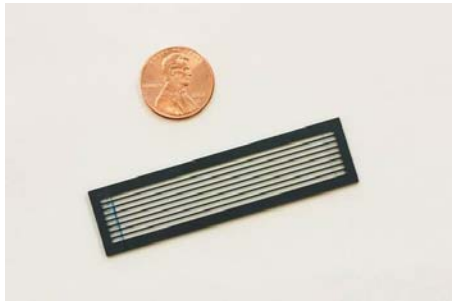
- DNA sequencing service
- Now with ABI 3730xl
- 96 sequences of 800 bases in 2 hours

⇒ *10 bases per second*

ABI 3730xl

2006: Next Generation

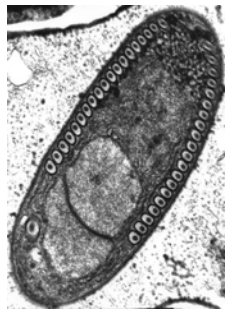
- Acquisition of the Solexa 1G system
- Q1 2007: 4 days to sequence a flow-cell
- One channel 500’000 reads of 26 bases
- 100 Mb per run



⇒ *300 bases per second*

Q2 2007: de novo Assembly of a 26 MB eukaryotic genome from less than 50 ng of genomic DNA

- *Microsporidia* are eukaryotes with genome of 20-30 Mb
- They are intracellular parasites of Daphnia
- DNA is very difficult to obtain
- Performed *de novo* assembly from single reads with EDENA
- Prof. Dieter Ebert and his group from the University of Basel are using this data to mine for candidate genes for host parasite interactions and for genetic markers (variable number tandem repeats VNTRs)
- Estimation of the genome size



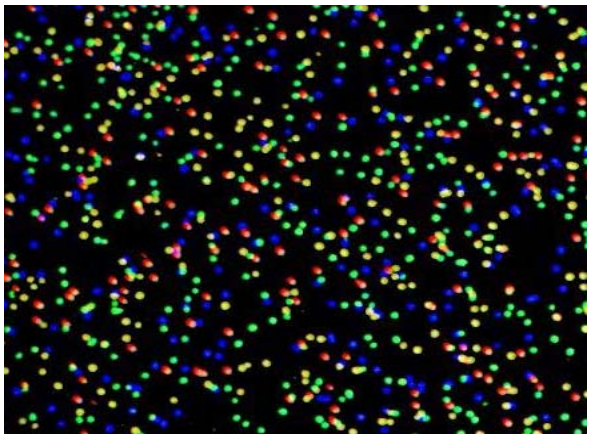
Microsporidia

Daphnia

End 2007: New recipes

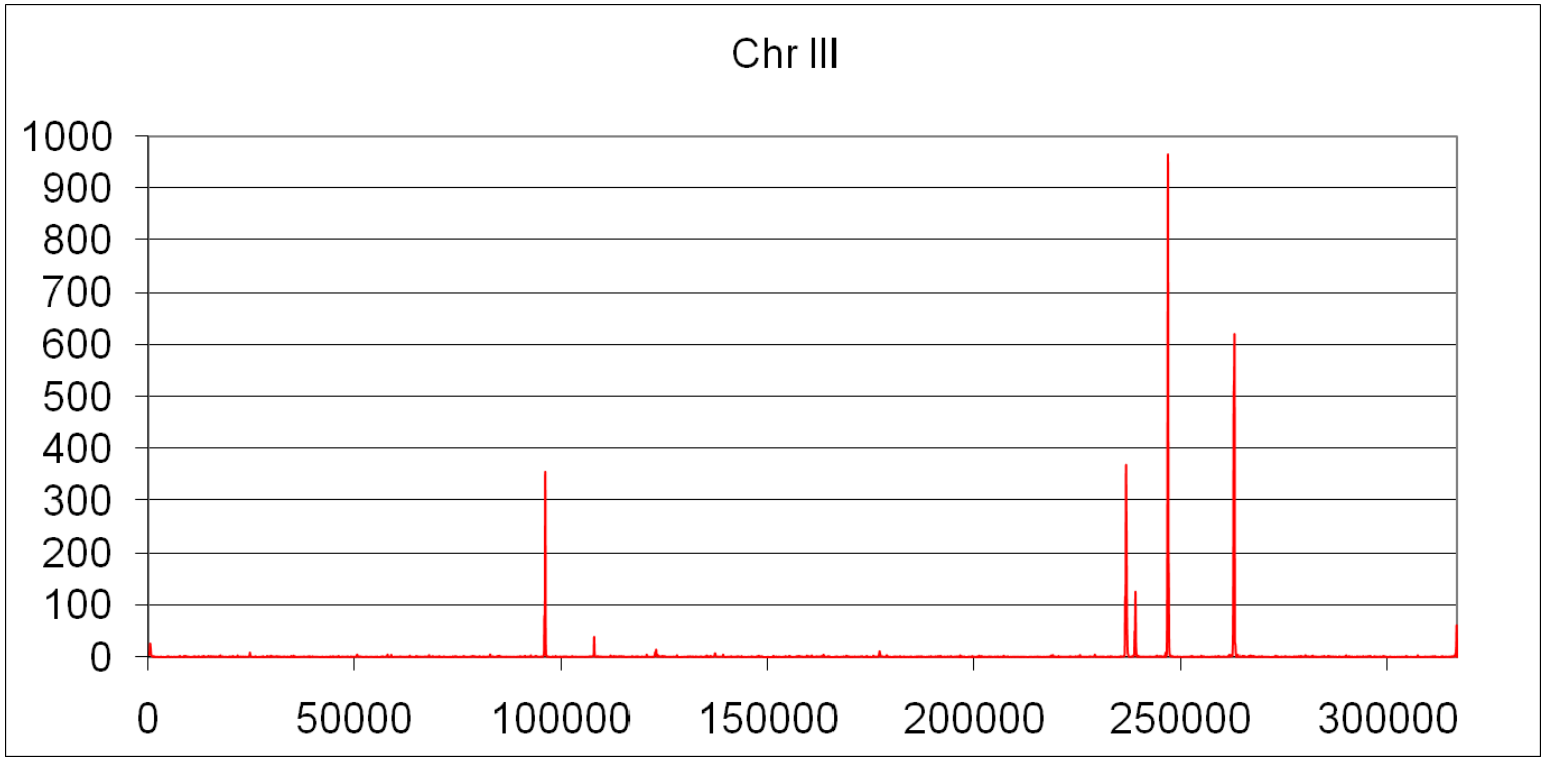
- More tiles per lane
- One channel 3-4 mio reads of 36 bases
- Run in 3 days
- 1.1 Gb per run

⇒ *4000 bases per second*



One image is 20’000 DNA Colonies

Representative map of 100 bp bin counts for Tbf1-Myc ChIP DNA on Yeast Chr. III



David Shore
Cyril Ribeyre
Victoria Martin
Harri Lempiäinen

Felix Naef
Enrico Guarnera
Jacques Rougemont
Gregory Lefebvre

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

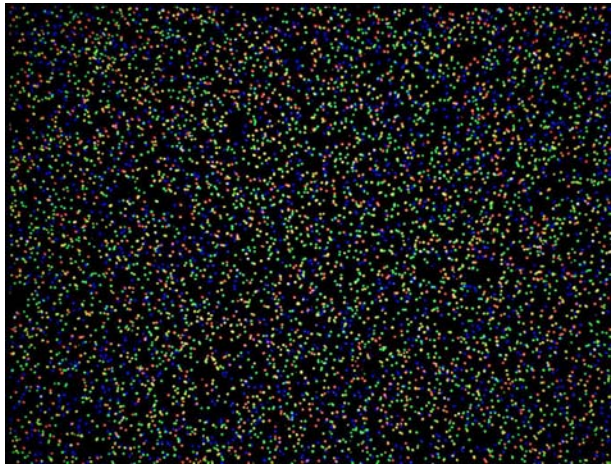
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

FASTERIS

2008: GAII

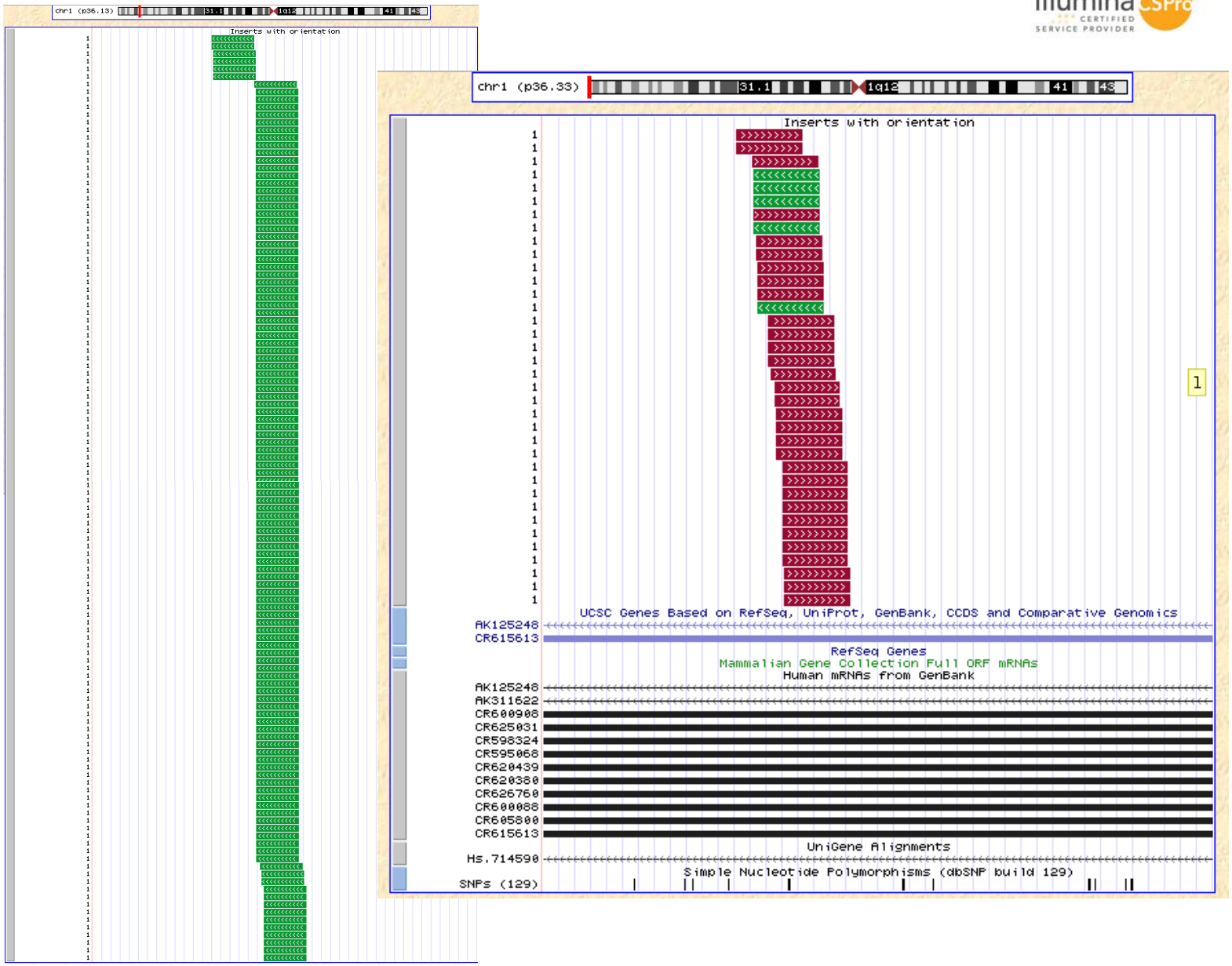
- Installation of the GAII upgrade
- Release of the paired-ends module
- New sequencing kits
- One channel 7'000'000 reads of 2x 36 bases
- 4 Gb per run in 5 days

⇒ *9'500 bases per second*



One image is 100'000 DNA Colonies

Small RNA mapping



2009: Pipeline 1.4

- GAIIx upgrade
- New kits
- New Pipeline 1.4
  - Higher cluster density
  - Real-time base-calling
- One channel 10-20'000'000 reads of 2x 76 bases
- up to 20 Gb per run in 8 days

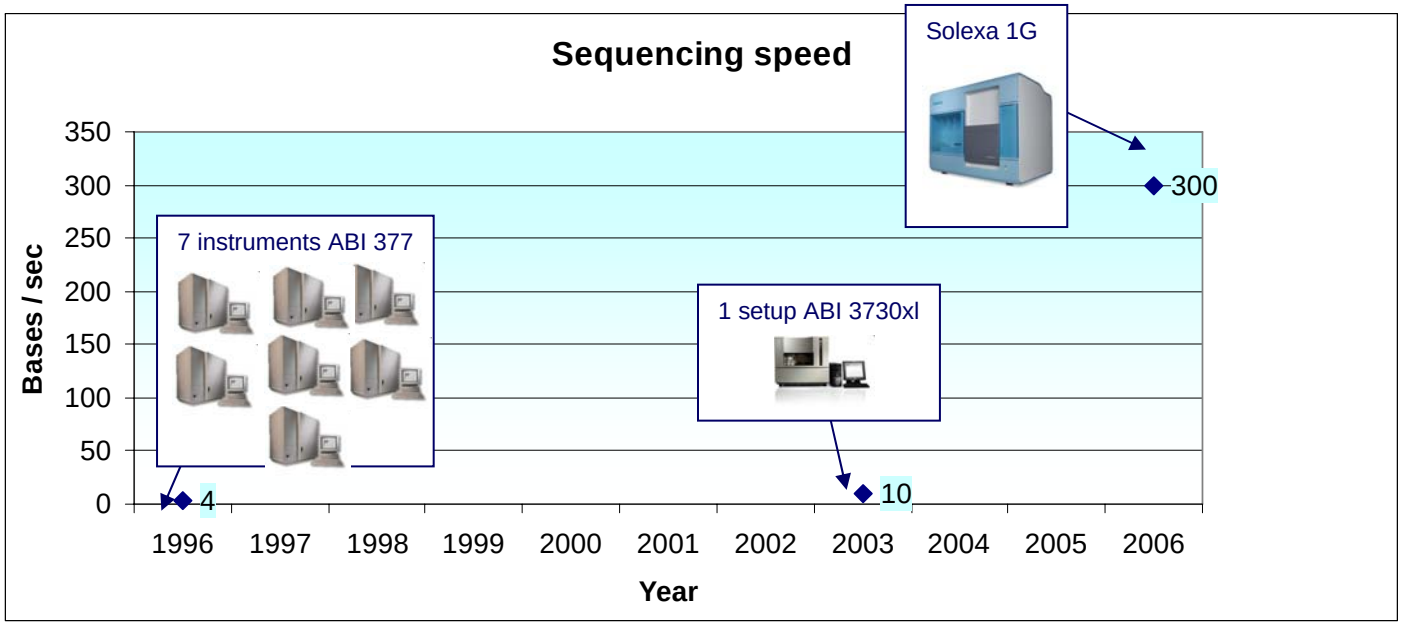
⇒ *35'000 bases per second*

One image is 220'000 DNA Colonies

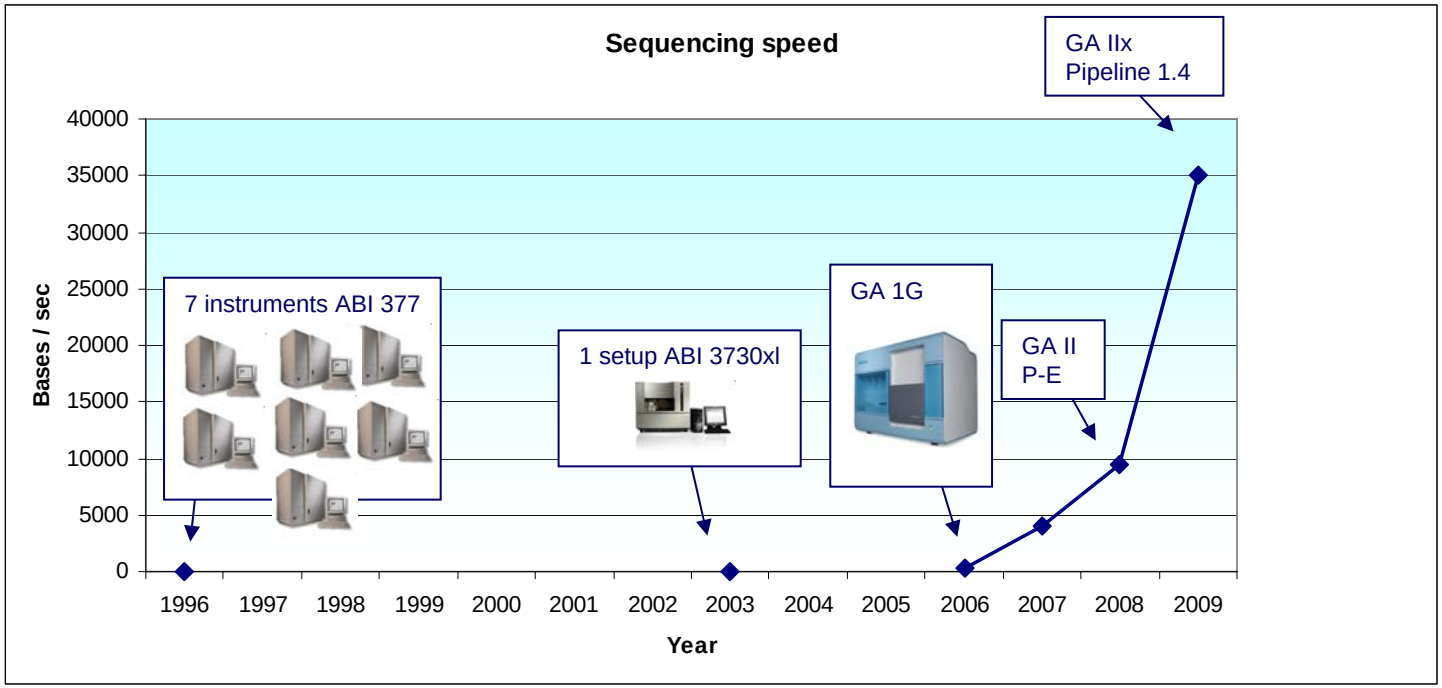
Fasteris Certified by illumina

- Fasteris and NCGR, USA, are the **first facilities** certified by illumina for **Genome Analyzer** Applications (Jan ’09)
- Illumina CSPro is a collaborative service provider partnership dedicated to ensuring the delivery of the highest quality data available for genetic analysis applications
- Illumina CSPros undergo a rigorous two-phase certification process that include minimum data generation, data certification and on-site audit of the facility and processes

The Genome Analyzer revolution



Tremendous Speed Increase



Throughput per channel

	Q1 2007	Q4 2007	Q2 2008	Q2 2009	2010
Read length	26	36	36	76	150
Paired	1	1	2	2	2
Millions of DNA Colonies	0.5	4	7	15	40
Millions of bases	**13**	**144**	**504**	**2'280**	**11'250**

Coverage

Bacterium	***3 Mb***	**4**	**48**	**168**	**760**	**3'750**
Arabidopsis	***120 Mb***	0.1	1.2	4.2	19.0	93.8
Human	***3 Gb***	0.0	0.0	0.2	0.8	3.8

High number of reads per channel Good for large genomes

- Not enough reads for large genome projects
  - => several channels or flow-cells still needed for mammalian genomes
- Soon 30x coverage in a single flow-cell

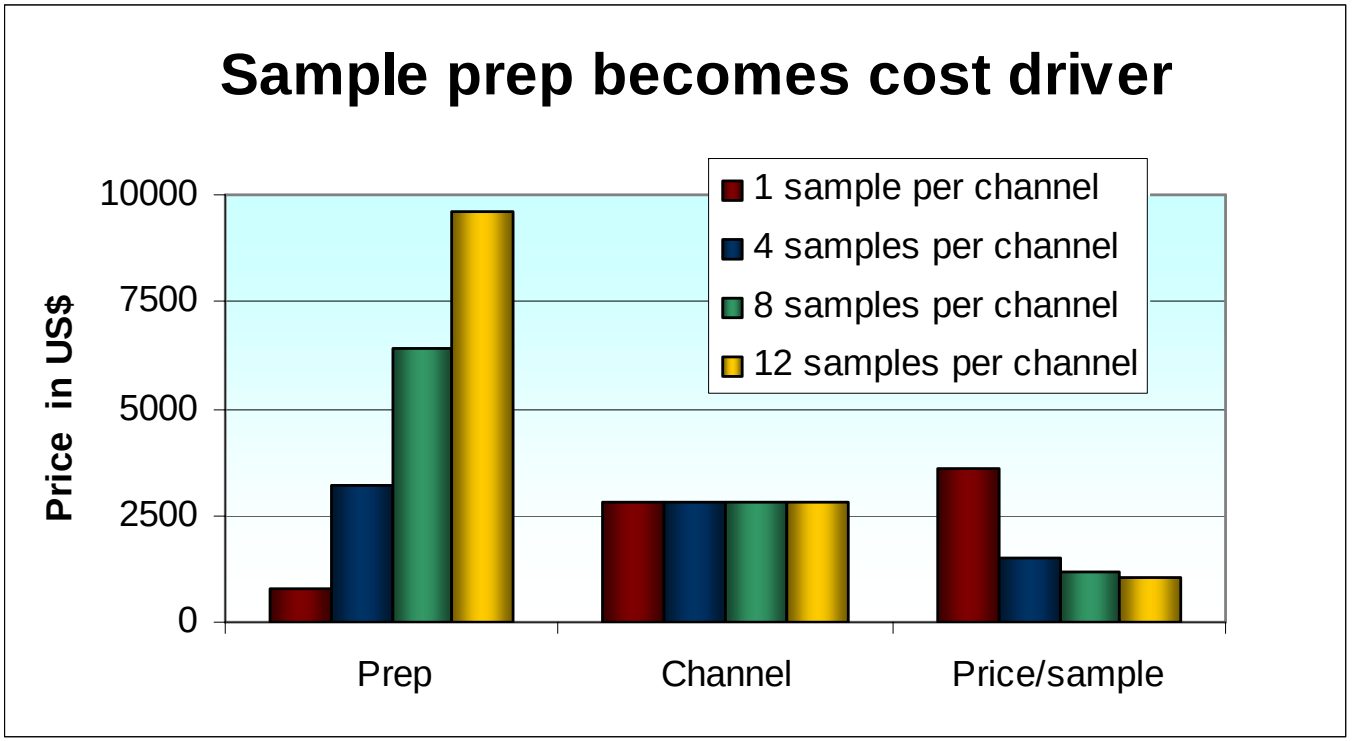
High number of reads per channel
Too many for small projects

- Too many reads for several applications

=> Need multiplexing

- 4, 8 or more samples per channel (ChIP-SEQ, smallRNAs, bacteria, targeted re-sequencing, etc..)

High number of reads per channel: Multiplexing



- Need for more efficient and cheaper sample prep

=> We are working on 96-wells preps

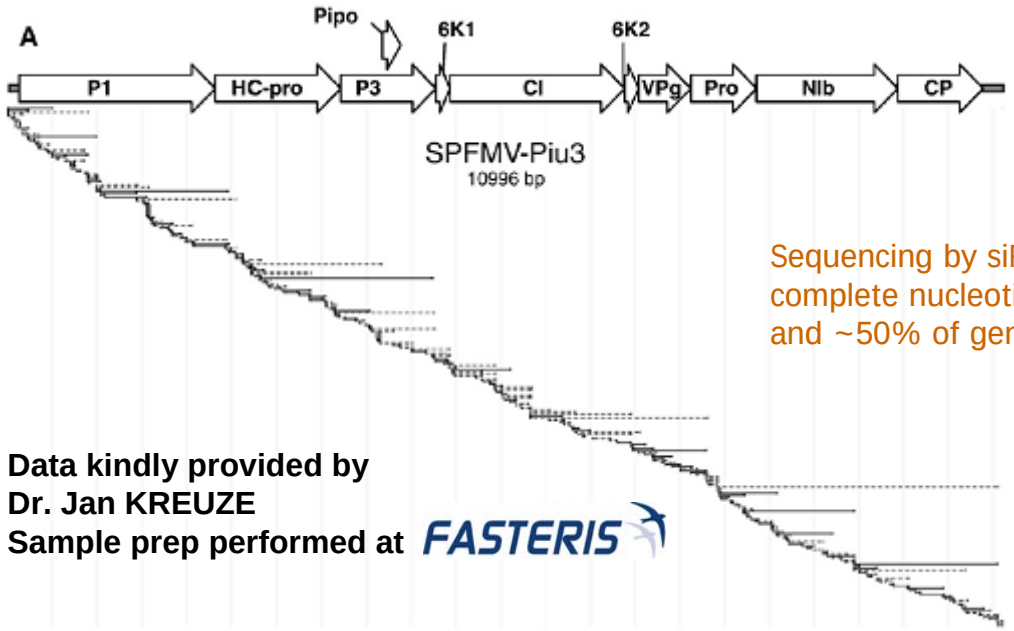
Dealing with longer reads

- Our standard runs are:
  - 1 x 38 bp
  - 1 x 76 bp
  - 2 x 38 bp
  - 2 x 76 bp
- It is becoming more and more difficult to fill the flow-cells and deliver results quickly
  - Without speaking of the other runs we do, e.g. 54 or 2 x 54 bp
- One 2 x 76 bp run takes 8-9 days
  - Only 2-3 runs per month

But new SBS enzyme and v7.0 recipes enable faster runs, *i.e.* 2 x 108 in 8-9 days

Here Come the Novel Applications

Using small RNAs to assemble *de novo* viral genomes



Sequencing by siRNA:
complete nucleotide sequence of SPFMV-Piu3
and ~50% of genome of three novel viruses

Data kindly provided by
Dr. Jan KREUZE
Sample prep performed at FASTERIS

Sequencing by siRNA: a novel generic tool for virus discovery

Kreuze et al. (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. ***Virology* 388: 1-7**

The Fasteris team at your service

Laurent Magne Anne Christelle Sylvie

Christine Cristel Elisabeth Cécile Marta Loïc