

***Recent developments
in small RNA analyses using
massively parallel sequencing***

Laurent FARINELLI

7 May 2010

Seminar: Next Generation Sequencing

Universität Bern, Switzerland





Outline

- ✈ A few words about Fasteris and the Genome Analyzer
- ✈ Small RNA sample preparation
- ✈ Bioinformatics analyses
- ✈ *de novo* assembly of small RNAs

Fasteris SA

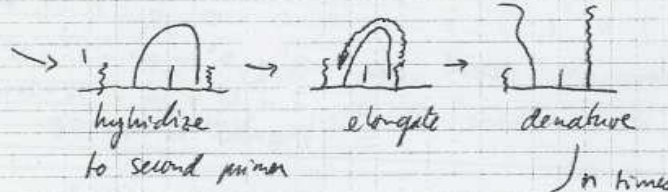
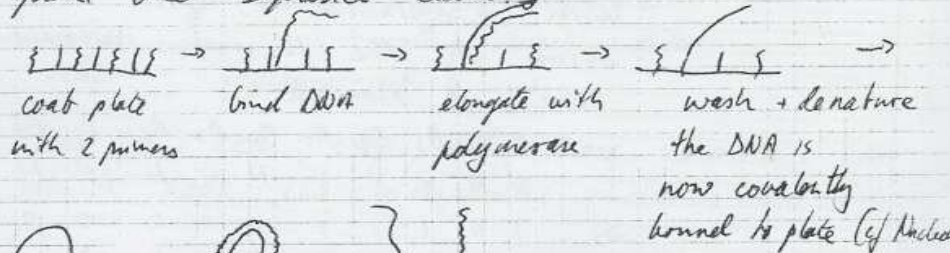
- ✈ Based in Geneva
- ✈ Founded 2003
 - Laurent FARINELLI and Magne OSTERAS
- ✈ Core business: DNA sequencing
- ✈ No external funding
- ✈ Importance of **SERVICE**



1996, invention of the DNA Colonies..

① PCR colonies (Pascal's idea)

- Coat a NucleoLink-like surface with 2 primers
- Apply diluted template DNA so that each molecule is $\approx 5 \mu m$ apart.
- Do a PCR reaction without primers in solution. The result should be spots of DNA amplified from one sequence each:



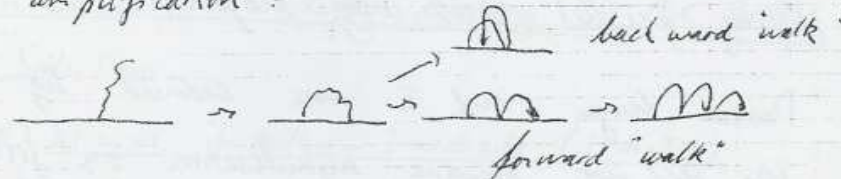
Signature: *Pavelis*

Read and understood: *SNV*

Date: 13.11.96

Date: 15.11.96

"Thus each spot of DNA will have been initiated with one molecule only, which "walked" during PCR amplification:



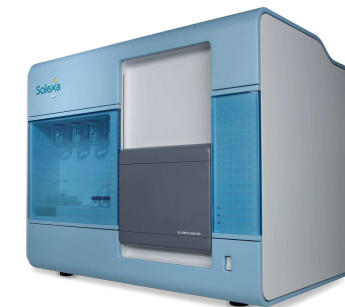
The resulting "PCR colonies" would be the equivalent to the coated beads.

1996-1997:

GlaxoWellcome's
Geneva Biomedical
Research Institute

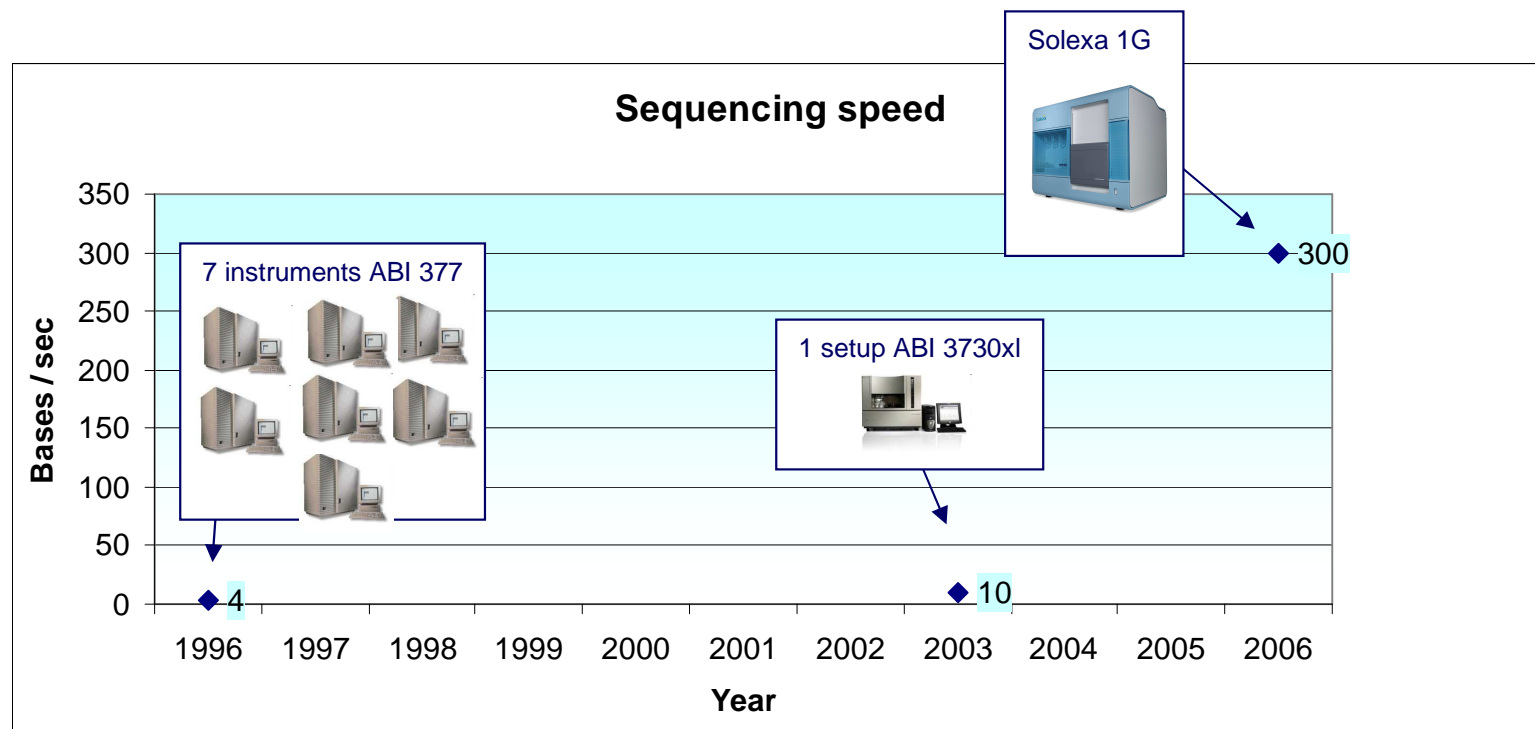
Mayer P., Farinelli L.
and Kawashima, E.,
1997, Patent application
WO 98/44151

**Key technology
for the illumina
Genome Analyzer**

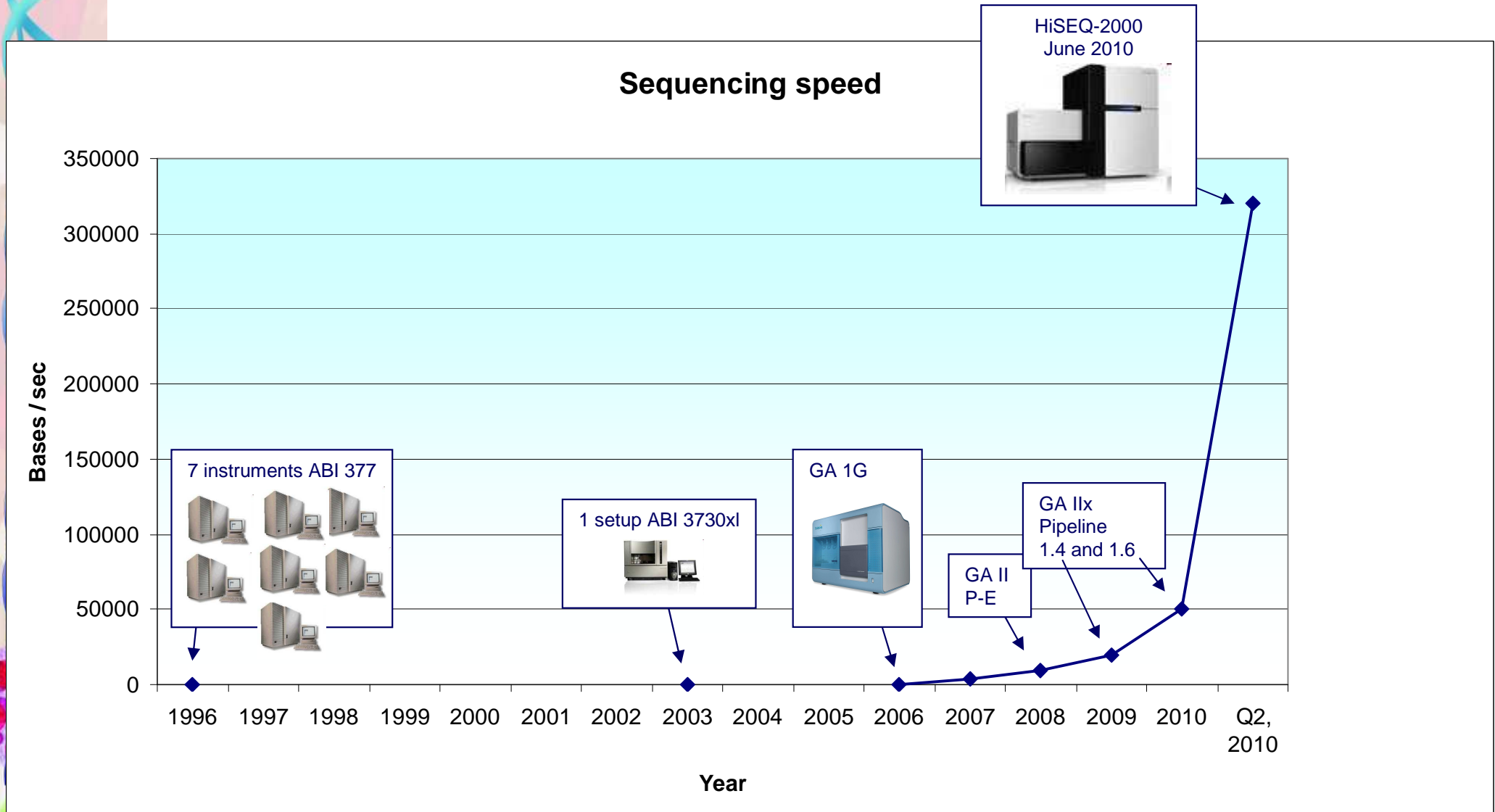


...now know as DNA Clusters

The Genome Analyzer revolution



Tremendous Speed Increase

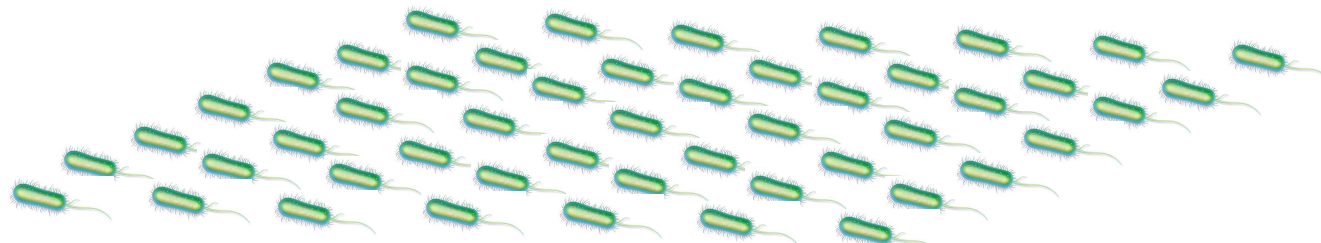


Today:

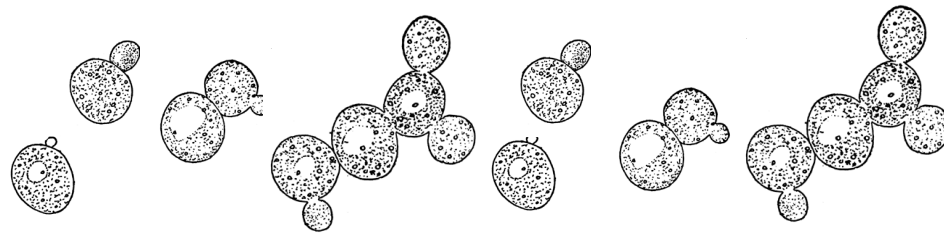
One GAIIX Paired-ends Channel

(One Run is 8 channels)

- 20-30 mio sequences of 2 x 108 bases
- 5.4 Gb
- 5'400 books of 1000 pages



- 1000x coverage of a 5 Mb bacterium



- 450x coverage of a 12 Mb Yeast

- 42x coverage of 128 Mb Arabidopsis

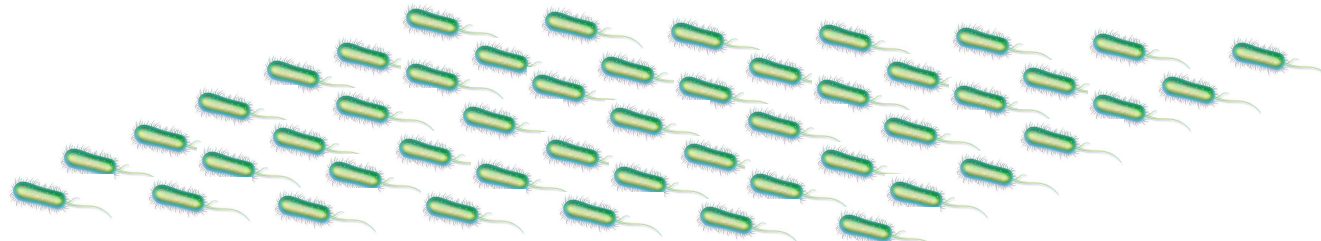


June 2010:

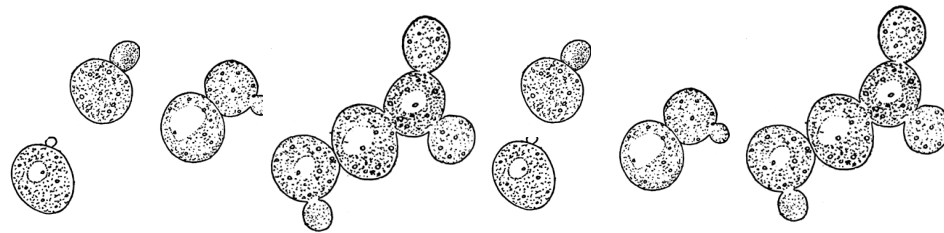
One HiSEQ 2000 Paired-ends Channel

(Illumina data: One 8-days run is 2 flow-cells of 8 channels each)

- 60 mio sequences of 2 x 108 bases
- 12 Gb
- 12'000 books of 1000 pages



- 2400x coverage of a 5 Mb bacterium



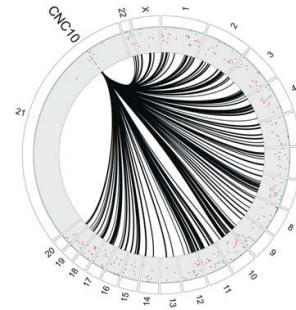
- 1000x coverage of a 12 Mb Yeast

- 93x coverage of 128 Mb Arabidopsis



Applications performed at FASTERIS

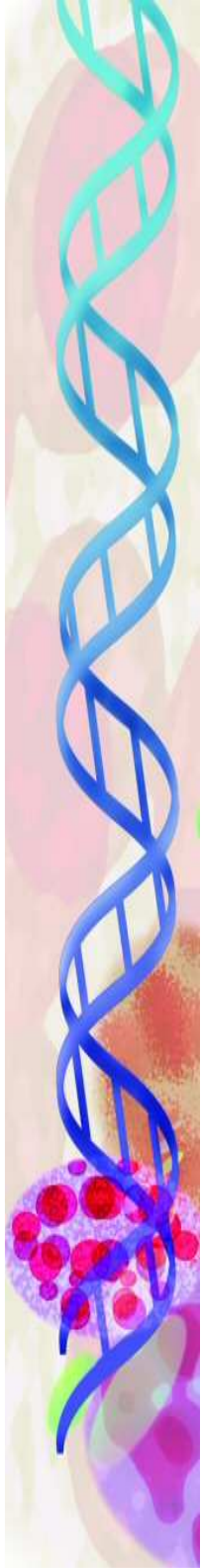
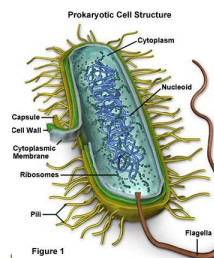
- Binding sites
 - ChIP-SEQ, 4C, etc..
- Small RNAs
 - Profiles, comparisons
 - de novo assembly
- Transcriptomes
 - Digital Gene Expression Profiling, mRNA-SEQ Whole
- Genomes
 - Re-sequencing
 - Bacteria, yeast, plant, animals, viruses
 - SNP detection
 - Targeted re-sequencing
 - PCR products, pooled samples
 - SureSelect target enrichment
 - *De novo* sequencing
 - Bacteria, eukaryotes
 - Metagenomics
- Custom applications
- For all applications: **Bar-coding / Multiplexing**
- **96-well preps**



Data kindly provided by Dr. Daniel ROBYR, in the group of Prof. Stylianos ANTONARAKIS, University of Geneva

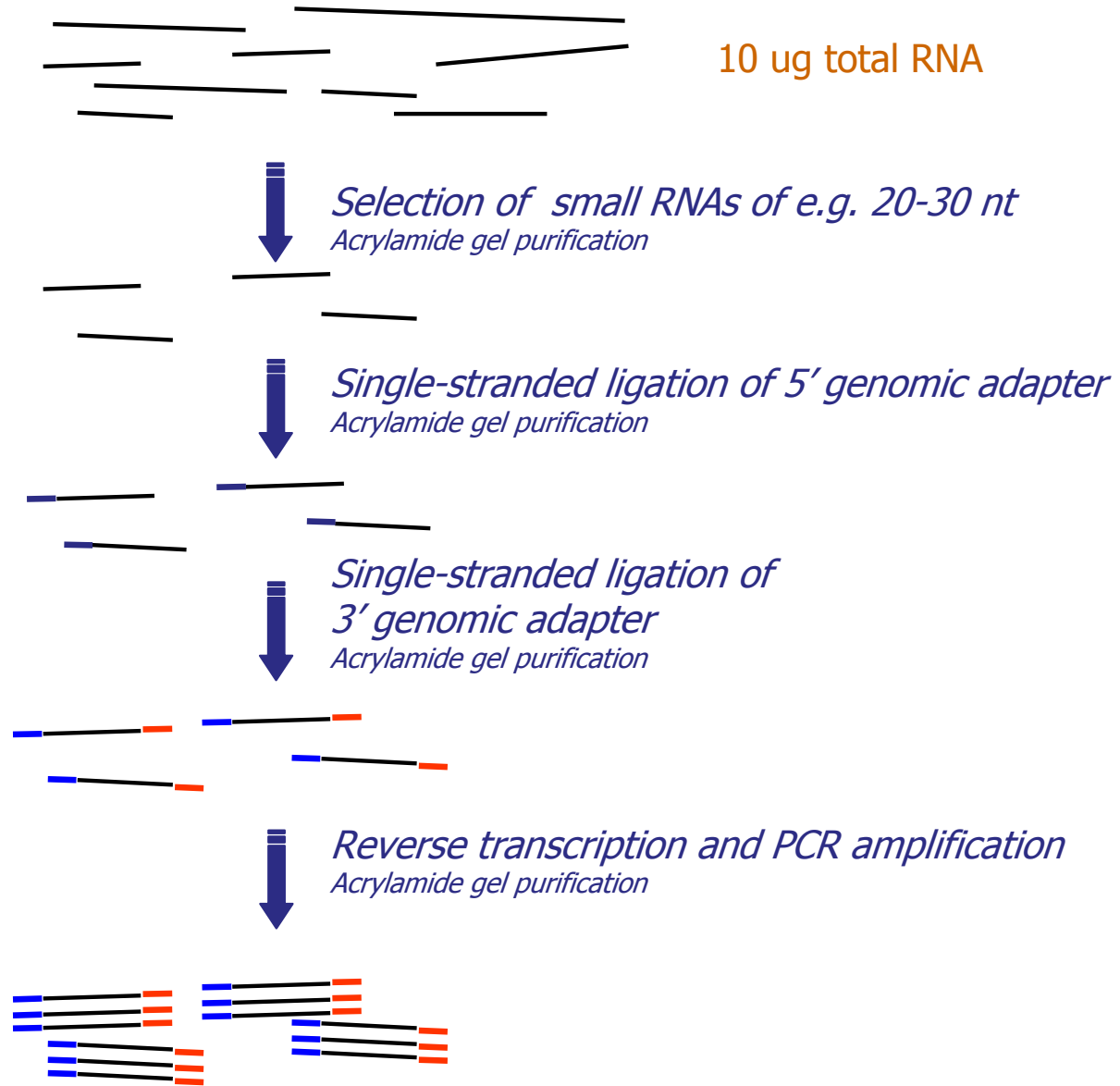


Agilent
Certified
Services Provider
Target Enrichment
System



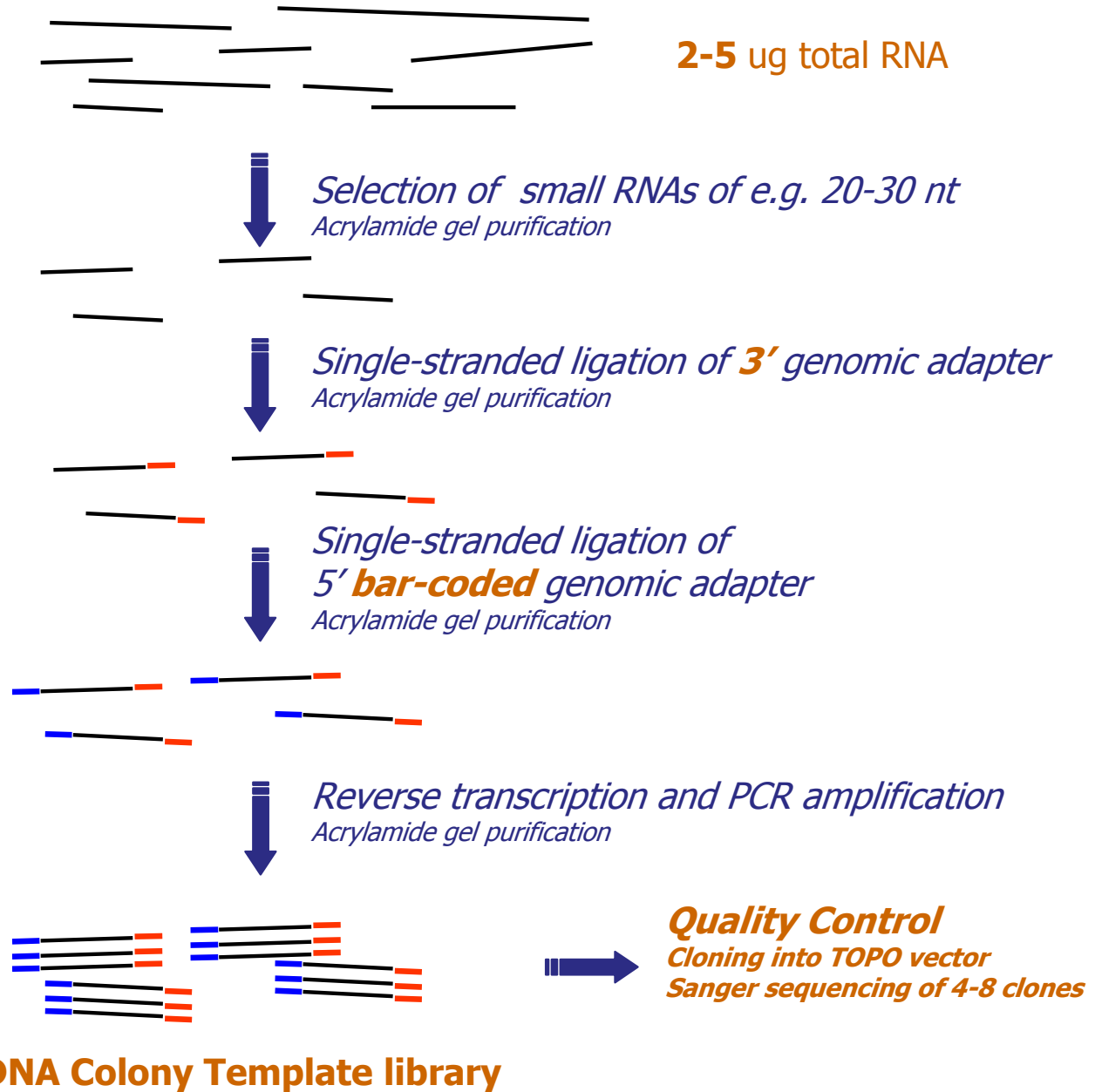
ILLUMINA Protocol v 1.0

small RNA Sample Preparation

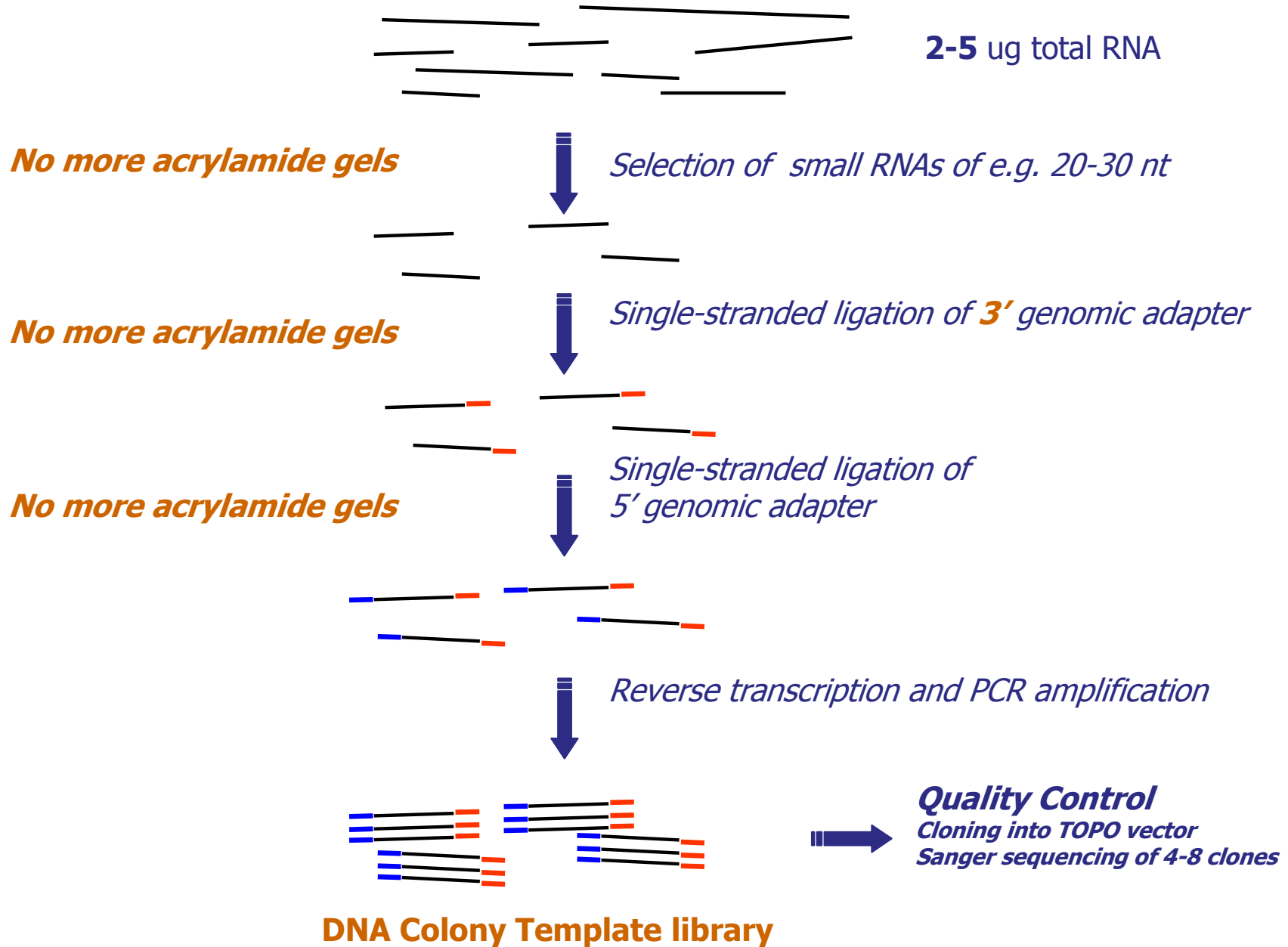


DNA Colony Template library

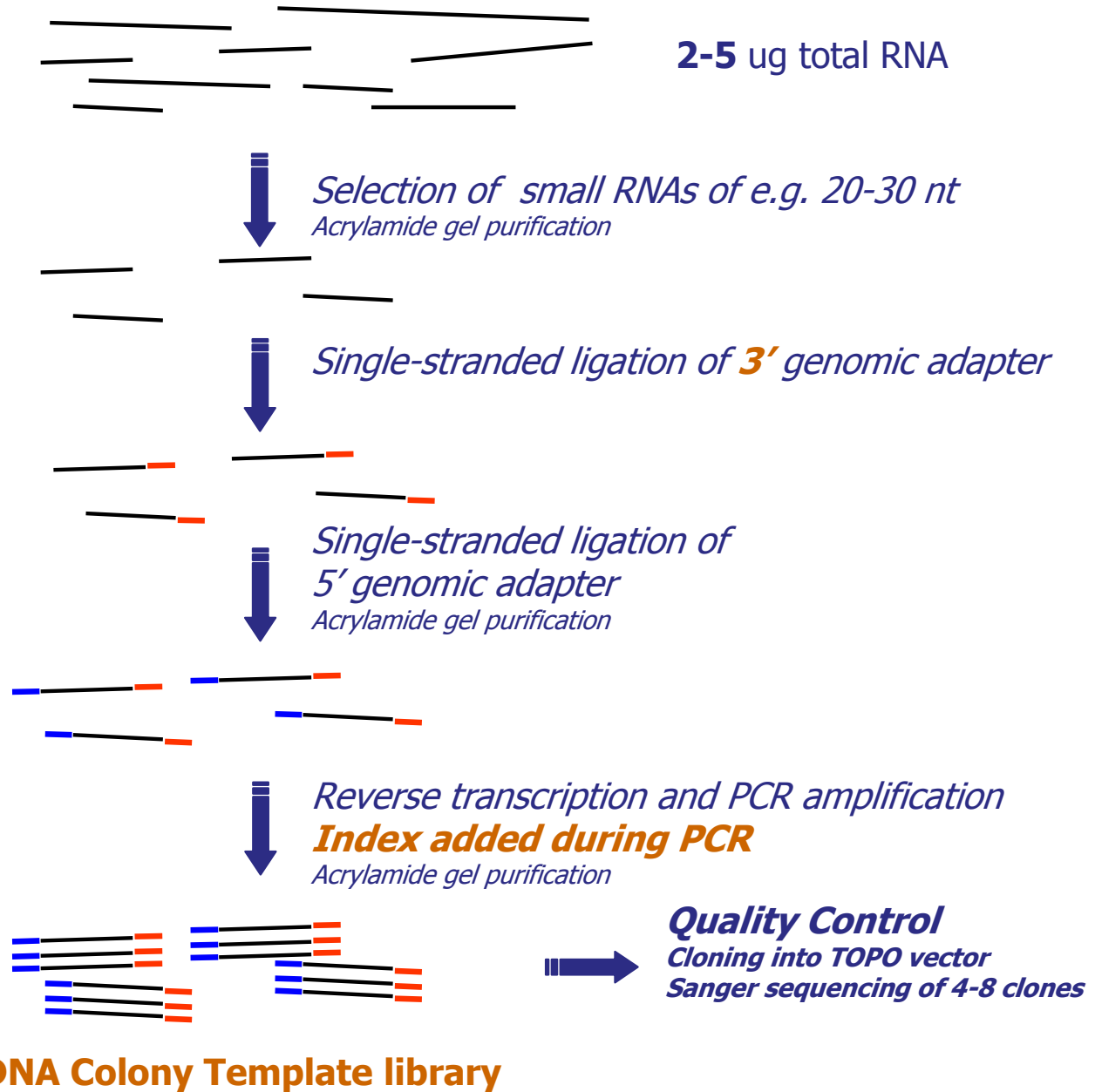
Fasteris modification of small RNA Sample Preparation



ILLUMINA protocol v 1.5 small RNA Sample Preparation

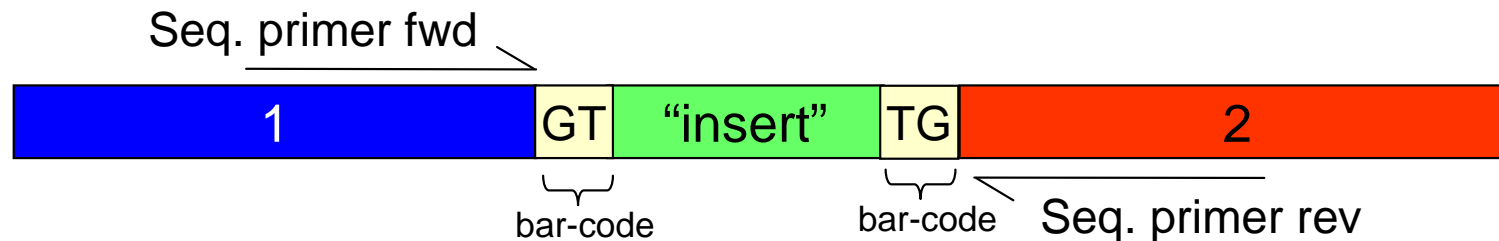


Fasteris modification of small RNA Sample Preparation



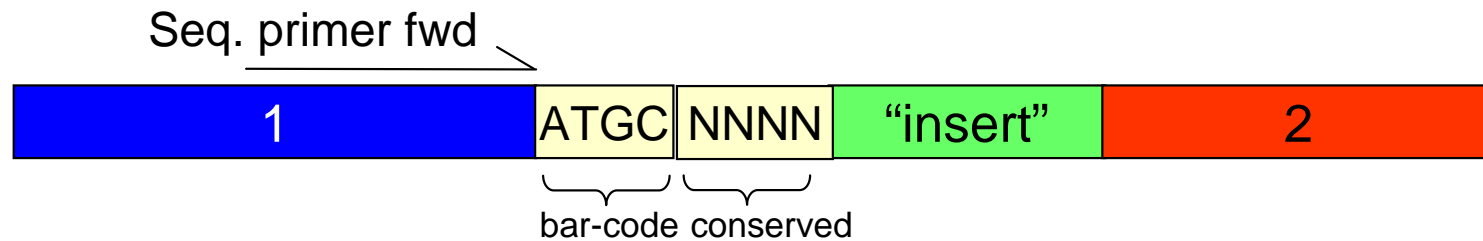
Multiplexing: "Fasteris" Bar-Codes

- To sequence several libraries in the same channel
- More important with increased number of reads per channel



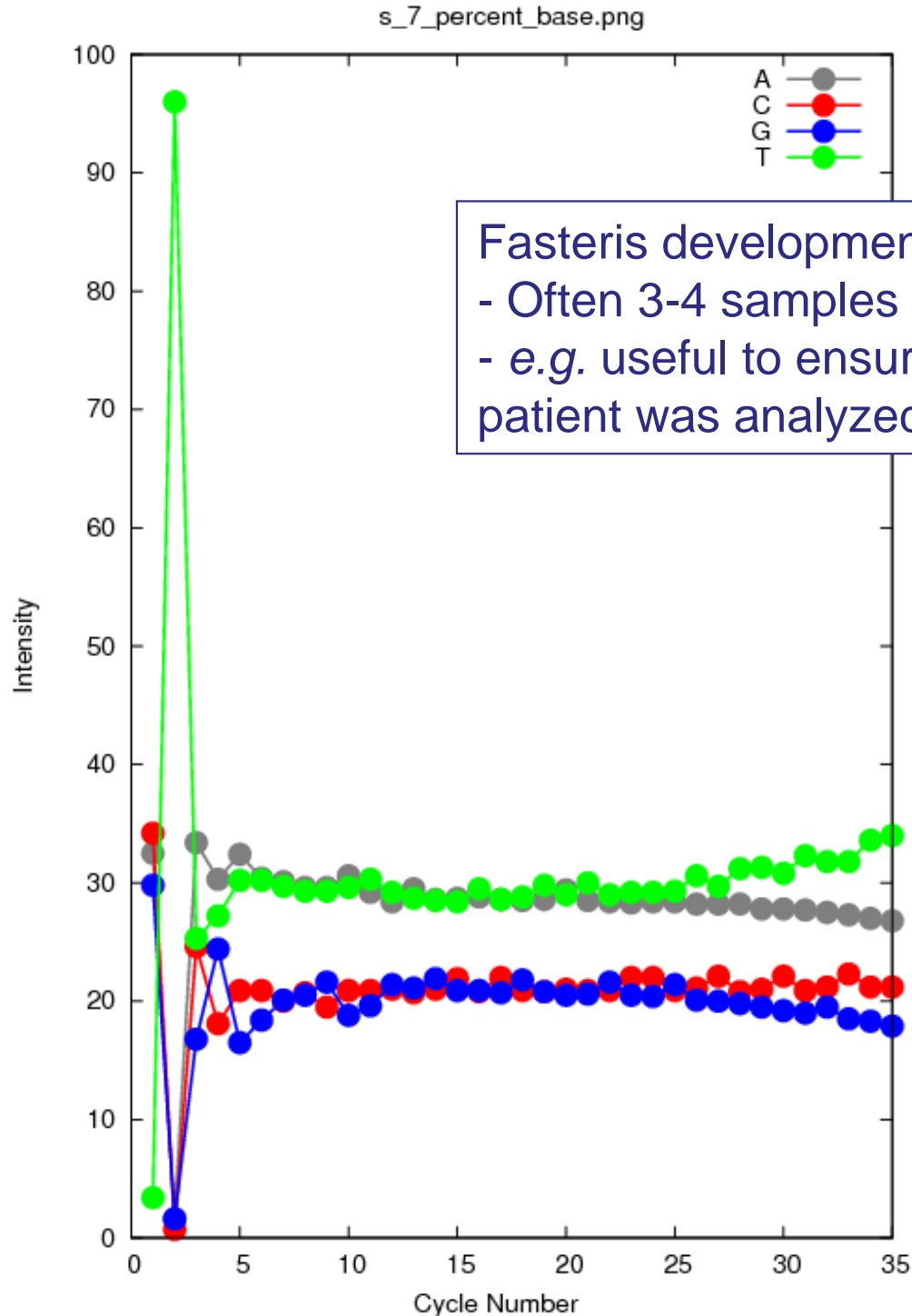
- Does not require paired-ends module, a single primer hybridization is sufficient
- Increased confidence: with paired-ends, both reads will start with the bar-code
- No additional cost if only one bar-coded channel in the flow-cell
- Sample prep: Must use bar-coded adapters at ligation step;
- Pooling can take place at this step when preparing large number of samples
- More attractive with longer reads
- For small RNA, bar-code on 5' adapter only;
- For GEX-NlaIII-MmeI, bar-code on adapter 2 only

Multiplexing small RNAs: "Fasteris" Bar-Codes



- Four bases bar-code on 5' adapter
- Common bases to avoid possible effect of sequence context during ligation

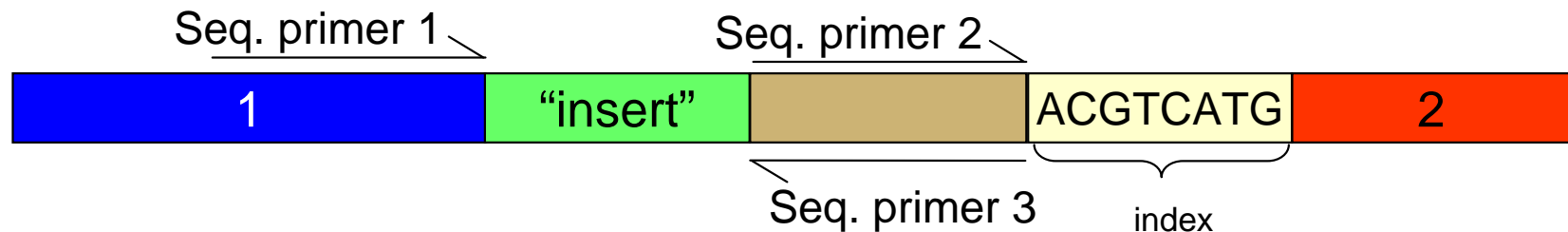
Bar-coding



Fasteris development
- Often 3-4 samples in one channel
- e.g. useful to ensure DNA from each patient was analyzed

Fasteris bar-codes, Dec. 2007

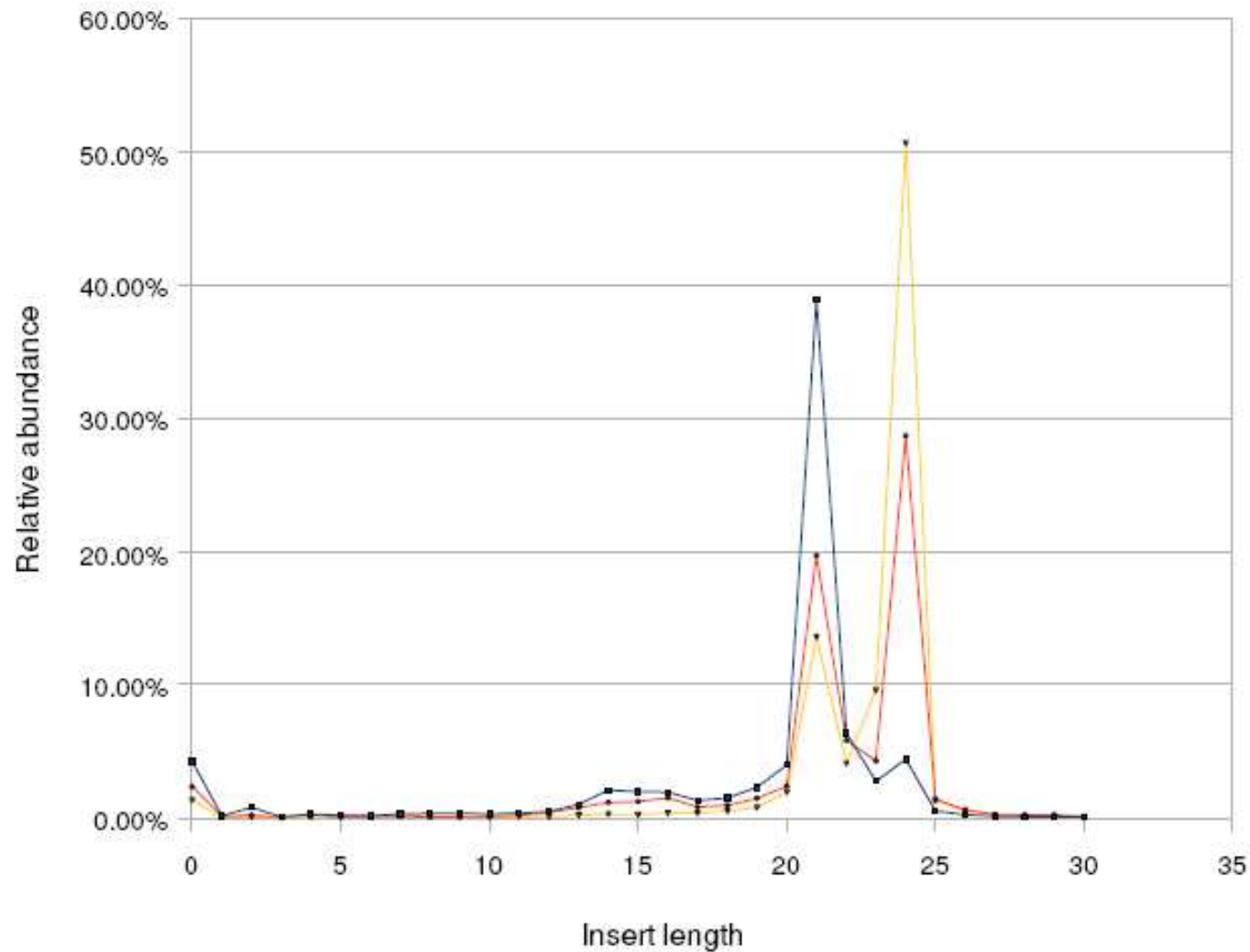
"Illumina Indexing"



- Needs a paired-ends module
- Must do ~8 more cycles, even if only one channel of the flow-cell with index
- Easier prep: Index can be added during PCR amplification
- Now also possible for small RNAs

small RNAs Profiles

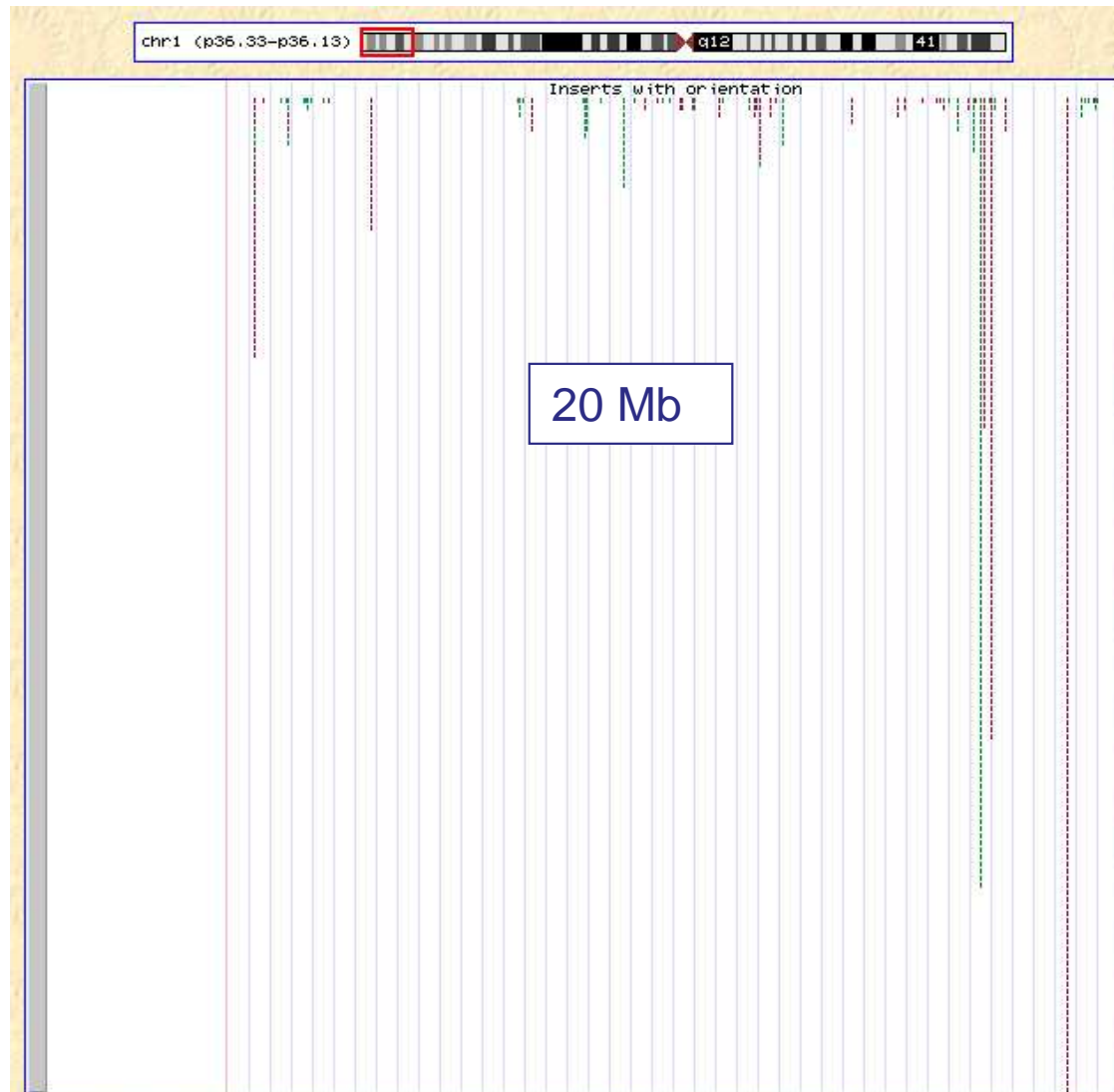
Insert profiles



Small RNA mapping



200 Mb



20 Mb

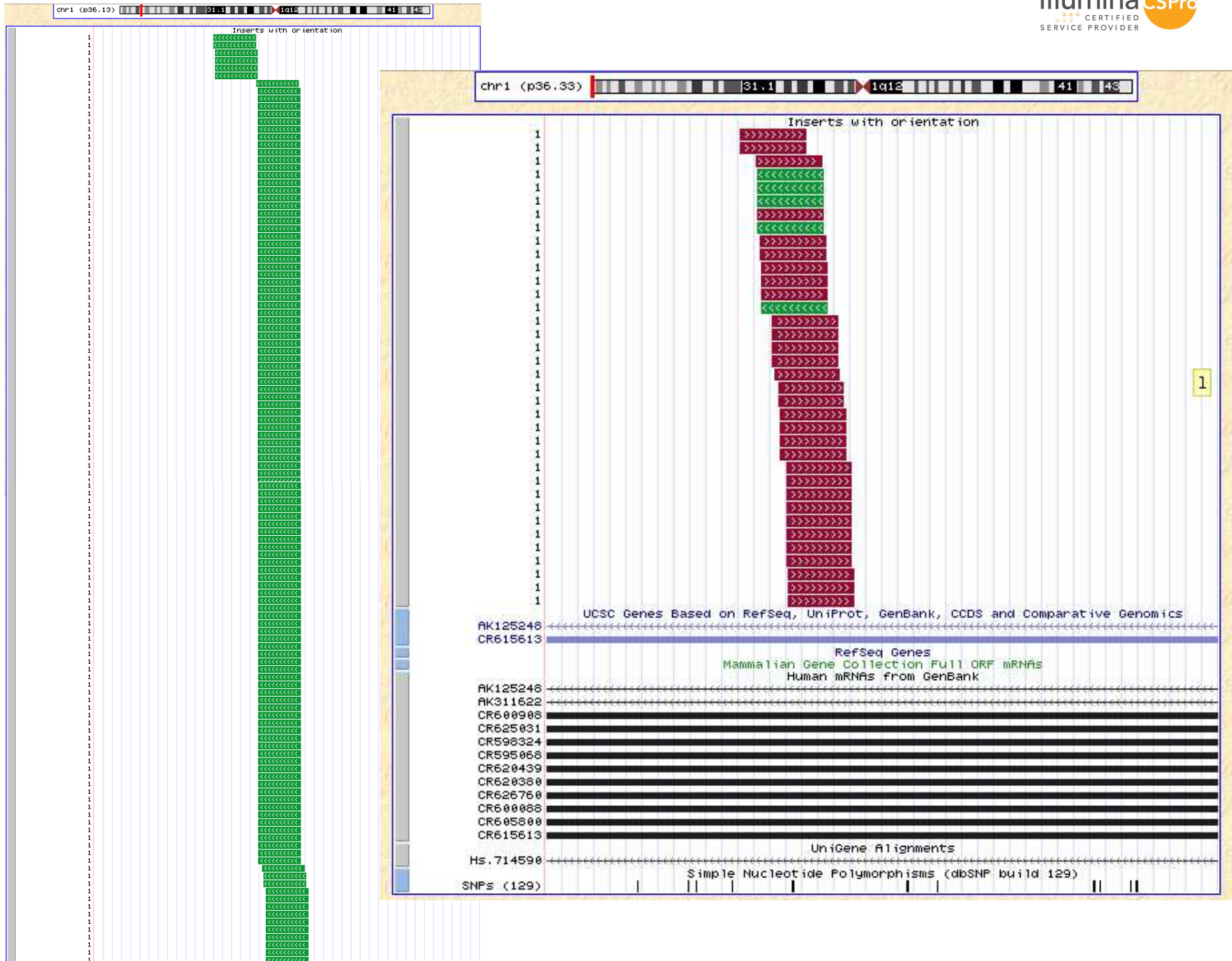


Small RNA mapping

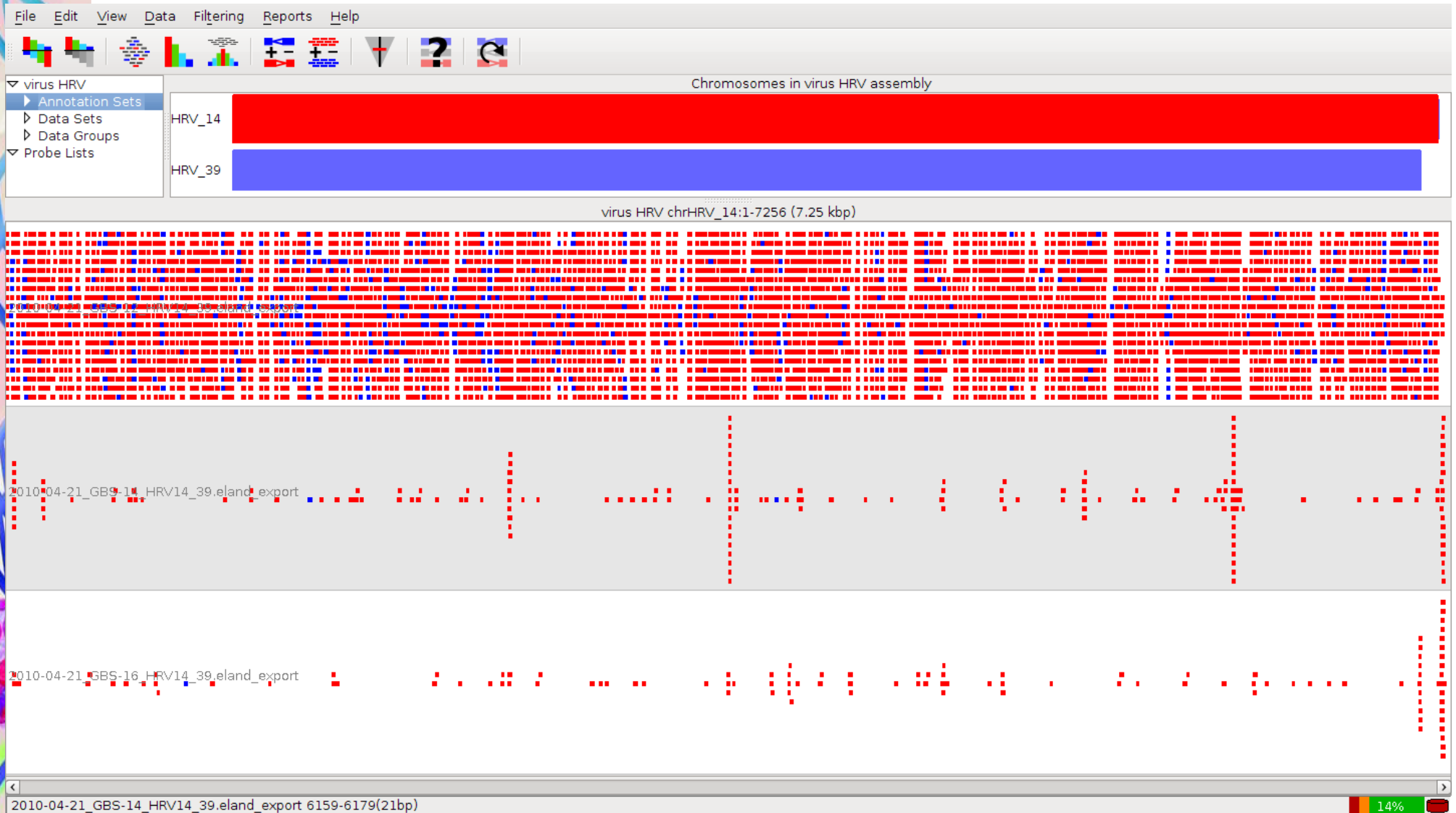


10 Kb

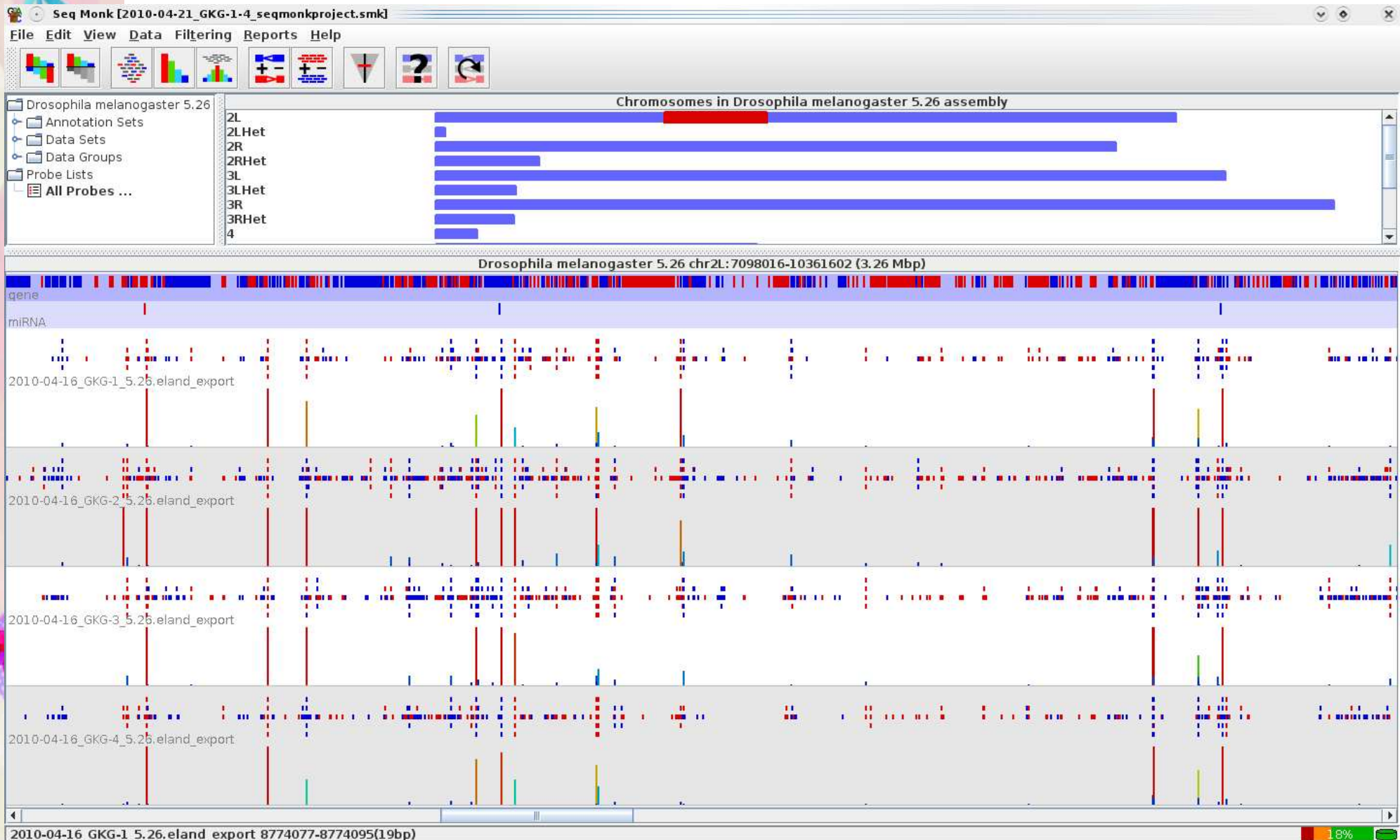
Small RNA mapping



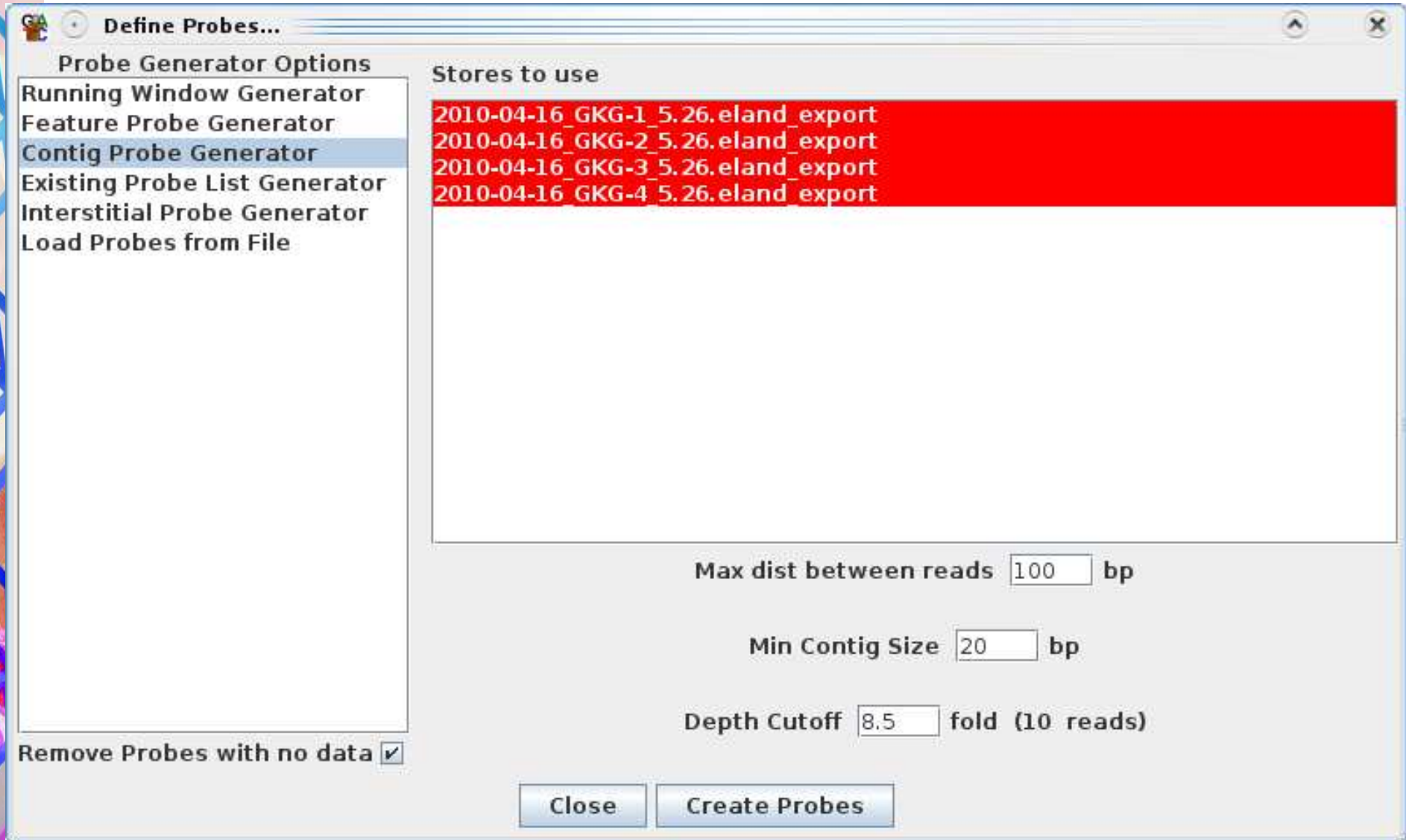
SeqMonk visualization tool



Visualization of miRNA using Seq Monk



Peaks detection of miRNA using Seq Monk



Define Probes...

Probe Generator Options

- Running Window Generator
- Feature Probe Generator
- Contig Probe Generator**
- Existing Probe List Generator
- Interstitial Probe Generator
- Load Probes from File

Stores to use

- 2010-04-16_GKG-1_5.26.eland_export
- 2010-04-16_GKG-2_5.26.eland_export
- 2010-04-16_GKG-3_5.26.eland_export
- 2010-04-16_GKG-4_5.26.eland_export

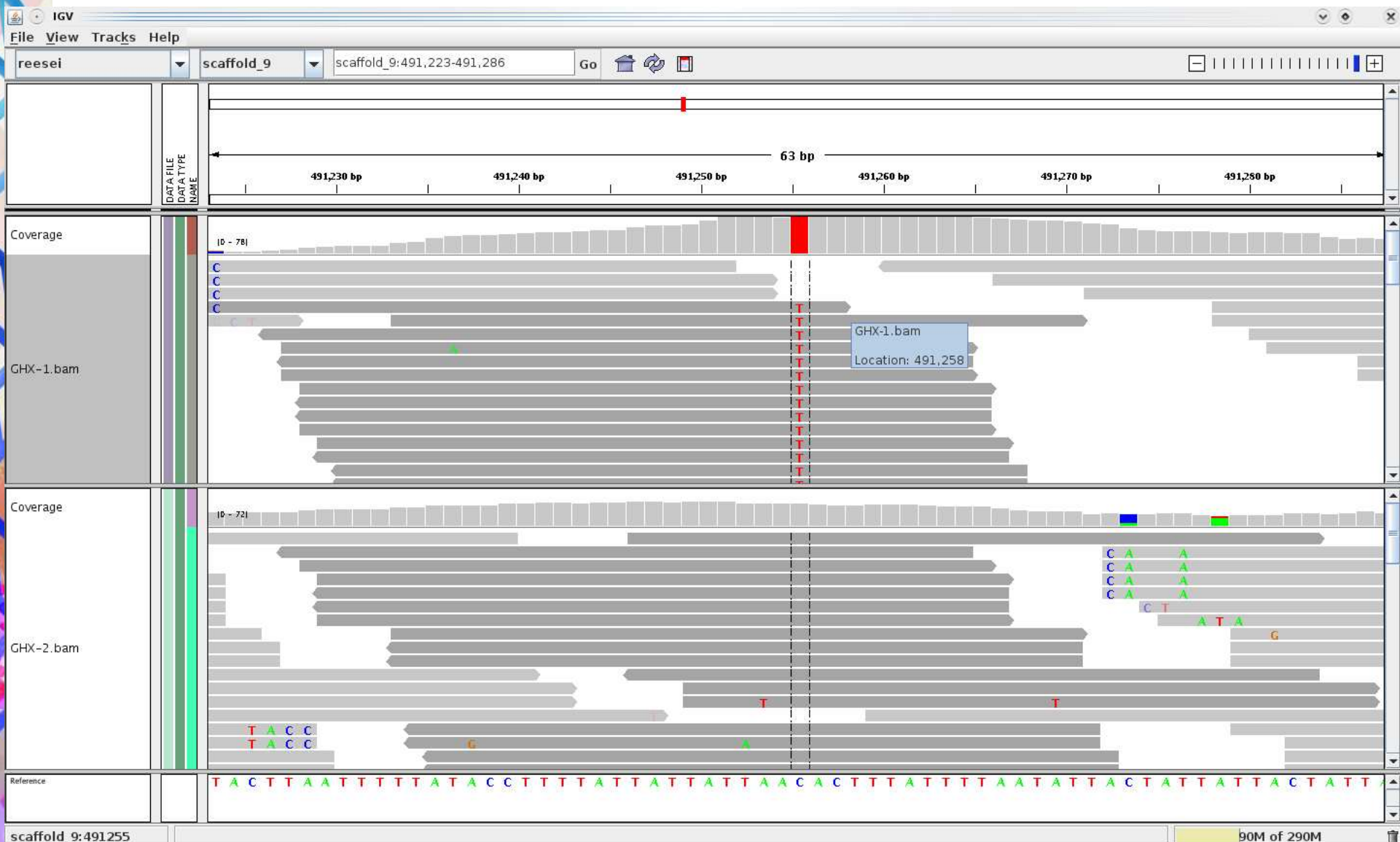
Max dist between reads bp

Min Contig Size bp

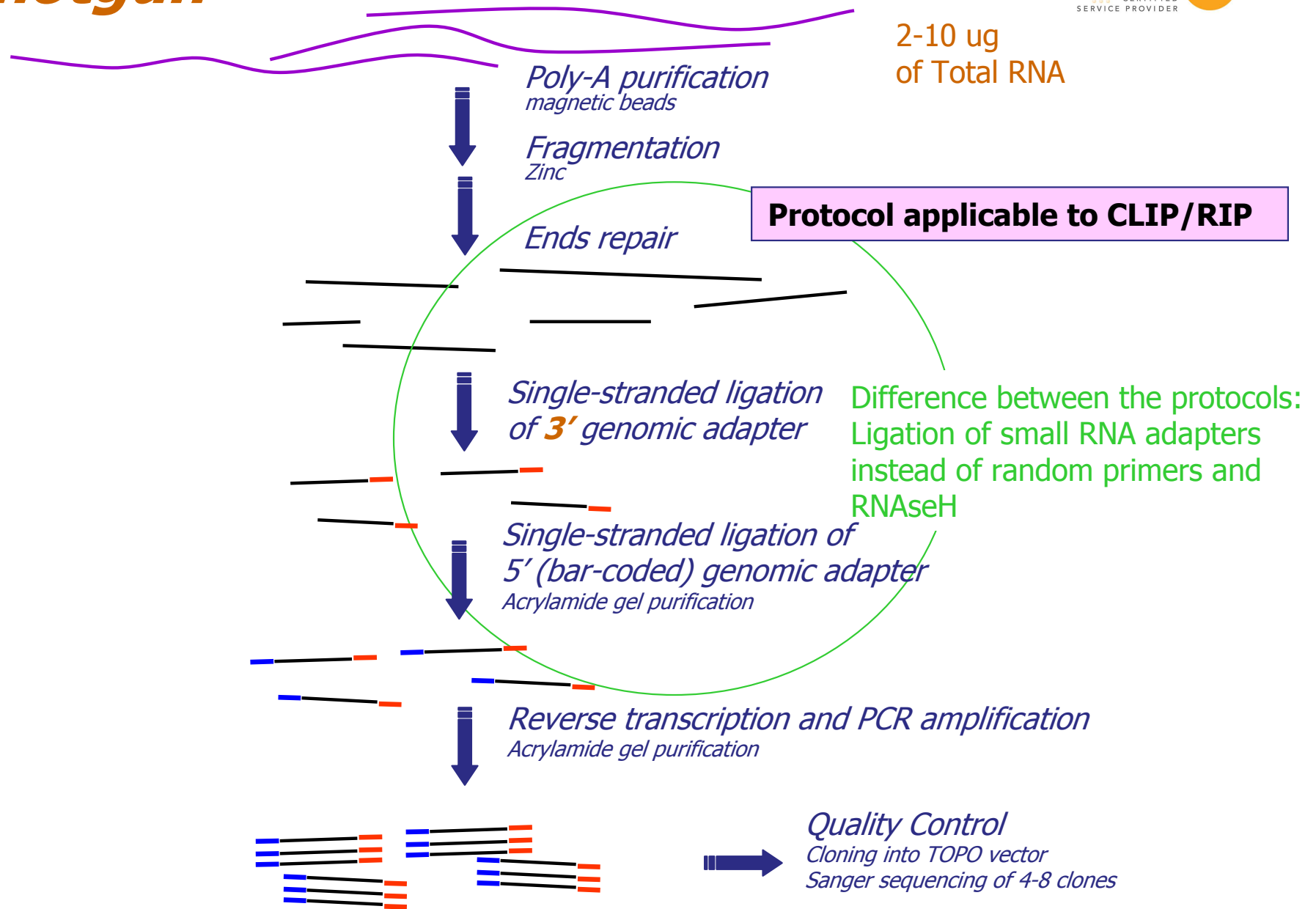
Depth Cutoff fold (10 reads)

Remove Probes with no data

IGV visualization tool for SNPs



"dir-mRNA-SEQ" Transcriptome Shotgun

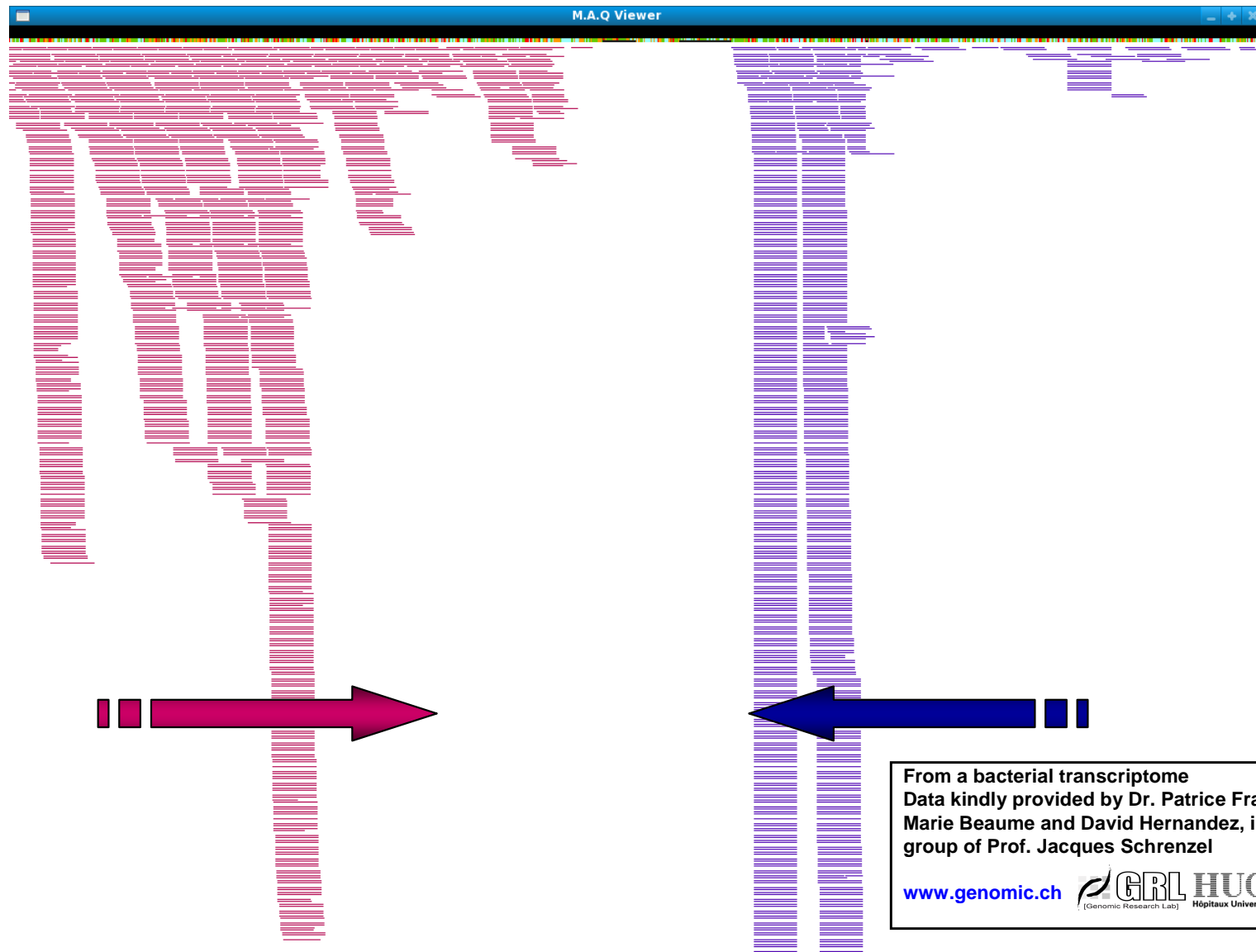


DNA Colony Template library

dir-mRNA-SEQ: Almost all reads are in the same orientation



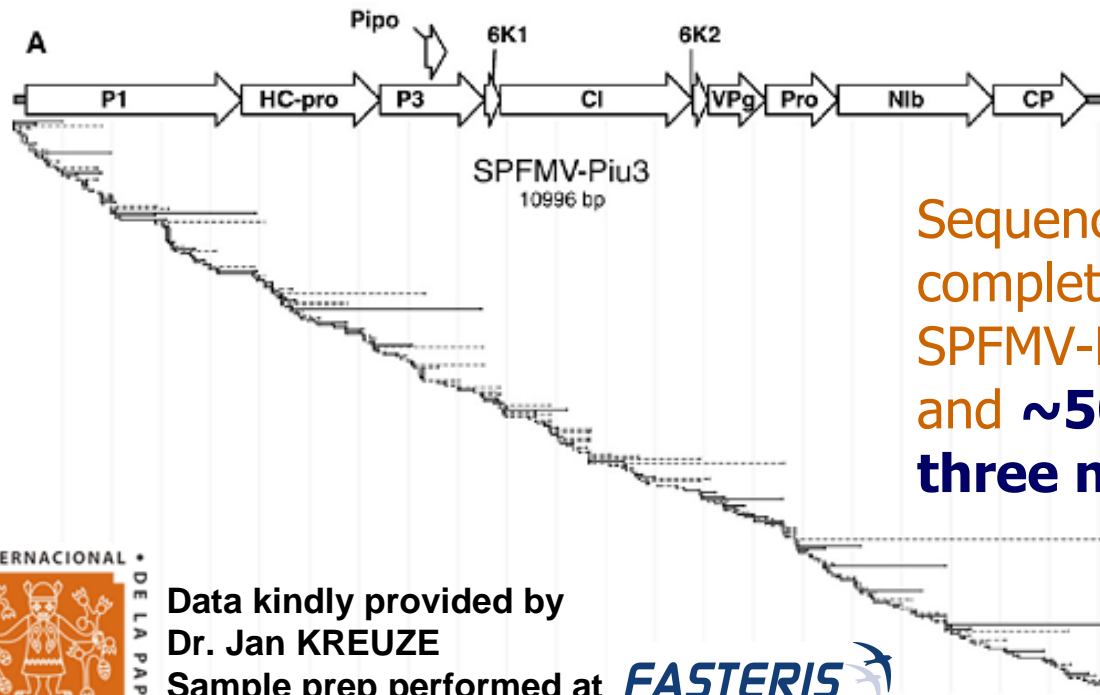
dir-mRNA-SEQ: Two transcripts in reverse orientation



From a bacterial transcriptome
 Data kindly provided by Dr. Patrice François,
 Marie Beume and David Hernandez, in the
 group of Prof. Jacques Schrenzel

www.genomic.ch  Hôpitaux Universitaires de Genève

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

De novo assembly of small RNAs from plants infected with virus

Kreuze approach:

- ✈ smallRNAs from infected sweetpotato plants.
- ✈ Assembled the smallRNAs into contigs using three assembling programs.
- ✈ Longest contig: 3115 nts.
- ✈ In one sample (dually infected), the contigs span the entire genome

Fasteris approach:

- ✈ smallRNAs from a control and an infected sample.
- ✈ Assembled the smallRNAs of the control sample into contigs.
- ✈ Mapped the reads of the infected sample on the obtained contig.
- ✈ Subtracted the mapped reads from the infected sample.
- ✈ Assembled the set of remaining reads (virus-specific reads) from the infected sample.

Results of the de novo assembling of the "virus-specific" reads

Velvet kmer size	17	17	17
# reads expected	85000	200000	500000
Sum of contigs length	22055	10885	6497
Number of contigs	106	39	16
N50 of contigs	210	310	1109
Average length of contigs	208	279	406
Maximum size	2658	2658	2841
Number of mapped reads (unique and multiples)	1031743/5221734 =(19.76%)	1056504/5221734 =(20.23%)	1016080/5221734 =(19.46%)

Blast of small RNA contigs

- ✈ Identical to plant ctDNA or mtDNA
 - Up to 3.5 kb !
- ✈ Identical to some virus sequences
- ✈ Identified 1000 new nt from unsequenced parts of the virus

UFRJ

Universidade Federal
do Rio de Janeiro

Collaboration with Prof. Maité VASLIN,
Federal University of Rio de Janeiro

small RNA contigs from the virus

- Alignment of the contigs onto know sequences from the virus

Total contigs 106

Aligned contigs 6

Total bases 22055

Aligned bases 3806 (17%)

- 6 of the *de novo* assembled match known virus sequences
- 85 contigs are from the plant
- 15 are not found in GeneBank, *e.g.* from plant or virus

UFRJ



Universidade Federal
do Rio de Janeiro

Collaboration with Prof. Maité VASLIN,
Federal University of Rio de Janeiro

The Fasteris team at your service



Laurent

Magne

Anne

Christelle

Sylvie

Christine

Cristel

Elisabeth

Cécile

Marta

Loïc