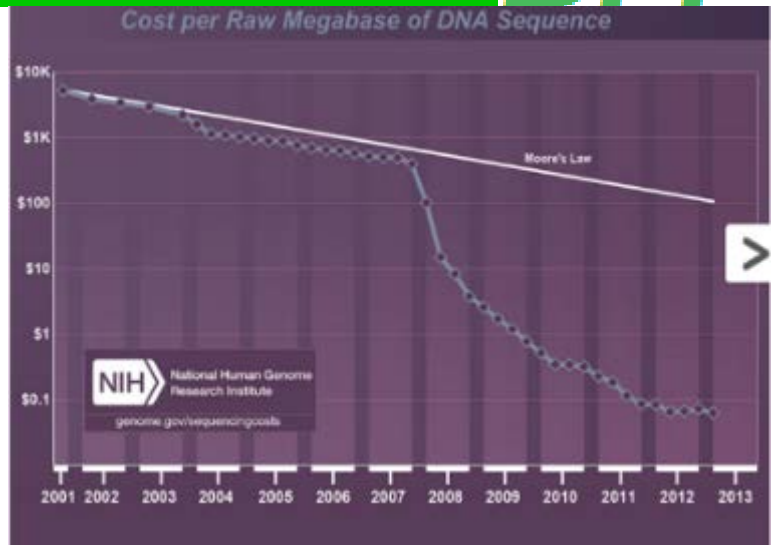
NGS-based viral diagnostics, let's confront some challenges

T. Candresse, A. Marais, C. Faure & S. Theil

thierry.candresse@inra.fr

Why such an impact of NGS ?

NGS



- Brings a **non-biased** and **potentially complete** vision of the sanitary state of a sample

- Speeds up the time between virus discovery and availability of a **diagnostic test**

- Allows to **improve classical PCR or LAMP assays** through the improvement of primers by integrating more complete information on viral diversity

- But also allows other studies, like virus **population genetics and evolution**, virus **ecology** ...

Diagnostics *vs* virus discovery

BFP

- **In most situations virus discovery imposes fewer constraints than diagnostics**
  (if no virus is found, then retest or test another sample…)
- **But the results of a diagnostic test should be as close as possible to the true infection status of a sample**
- **So a diagnostic test has to be**
  - Sensitive
  - Specific (or broad-spectrum)
  - Accurate (absence of false positives or of false negatives)
  - Repeatable
  - *and also cost- and time-effective*

Sensitivity of NGS-based approaches

BFP

- **Very few direct comparisons with PCR or ELISA**
- **Sensitivity known to be impacted by**
  - Viral concentration
  - Sequencing depth
  - Bioinformatic pipeline
- **COST proficiency test (siRNAs)**
  - Only **1 / 3rd** of pipelines had 100% sensitivity *(3 with false positives)*
  - …and only **2 / 3rd** had full repeatability… (2 pseudoreplicates)
- **Sensitivity possibly impacted by**
  - Target nucleic acid population
  - N.A. extraction/purification protocol (cf. host)
  - Mixed infections
- **Overall, testing currently less trivial than some might have hoped**

Mixed infections and coverage

BFP

Barley samples, dsRNASeq, multiplexed MiSeq 2x250 bp

Two constrated situations depending on the presence or not of the Barley endornavirus (dsRNA virus)

	total reads	Viral reads	Endornavirus	BaYMV RNA1	BaYMV RNA2	New virus RNA1	New virus RNA2
# 38	350,000	53%	**2000x**	**0.14x**	**2x**	**0.2x**	**3.9x**
# 15	283,822	59%	na	**140x**	**2000x**	**1600x**	**1000x**

- Strong variation in coverage limited ability to assemble the genomes of co-infecting viruses when the Endornavirus was present
- Could limit the ability to detect one of the viruses, in particular if lower enrichment or sequencing depth,
- Overall, **Excellent correlation with PCR detection (95.2%)**, two infection cases with <0.1% of reads but not detected by PCR

(Rolland et al., 2017, PLOS One, *in press*)

Specificity of NGS-based approaches

BFP

- **Specificity: critical for identification**
- **Because NGS diagnostics is sequence-based and unbiased, specificity should not be a concern**
- **But there are situations where data analysis may provide ambiguous results**
  - Novel virus: Blast analysis may not easily separate between presence of a novel agent or of a distant isolate of a known one (in particular if only partial genome coverage)
  - Closely related viruses. Blast and/or mapping analyses may not allow easily to know which virus is present (or both !)
  - In particular, for mapping there is a fine balance between too stringent parameters (may miss a divergent isolate) and too relaxed parameters (give a cross-mapping signal with a closely related virus)

Determining the infection status

BFP

- **Italian cherry sample, dsRNA sequencing**

	Contigs	% of total reads	Blast e-values
Little cherry virus 1	14	28.0%	0 to 6e-49
Apple chlorotic leaf spot virus	13	9.6%	0 to 4e-12
Prune dwarf virus	8	5.9%	0 to 1e-30
Cherry green ring mottle virus	2	0.5%	6e-64 to 3e-27
Peach mosaic virus	2	0.2%	6e-40 to 3e-30
Cherry necrotic rusty mottle virus	1	0.2%	2e-10
Potato virus T	1	16.0%	7e-68
Mint virus 2	1	11.1%	8e-86
Banana mild mosaic virus	1	2.7%	4e-11
Scaveola virus A	1	0.8%	3e-16

Which viruses infect the tested cherry tree ??

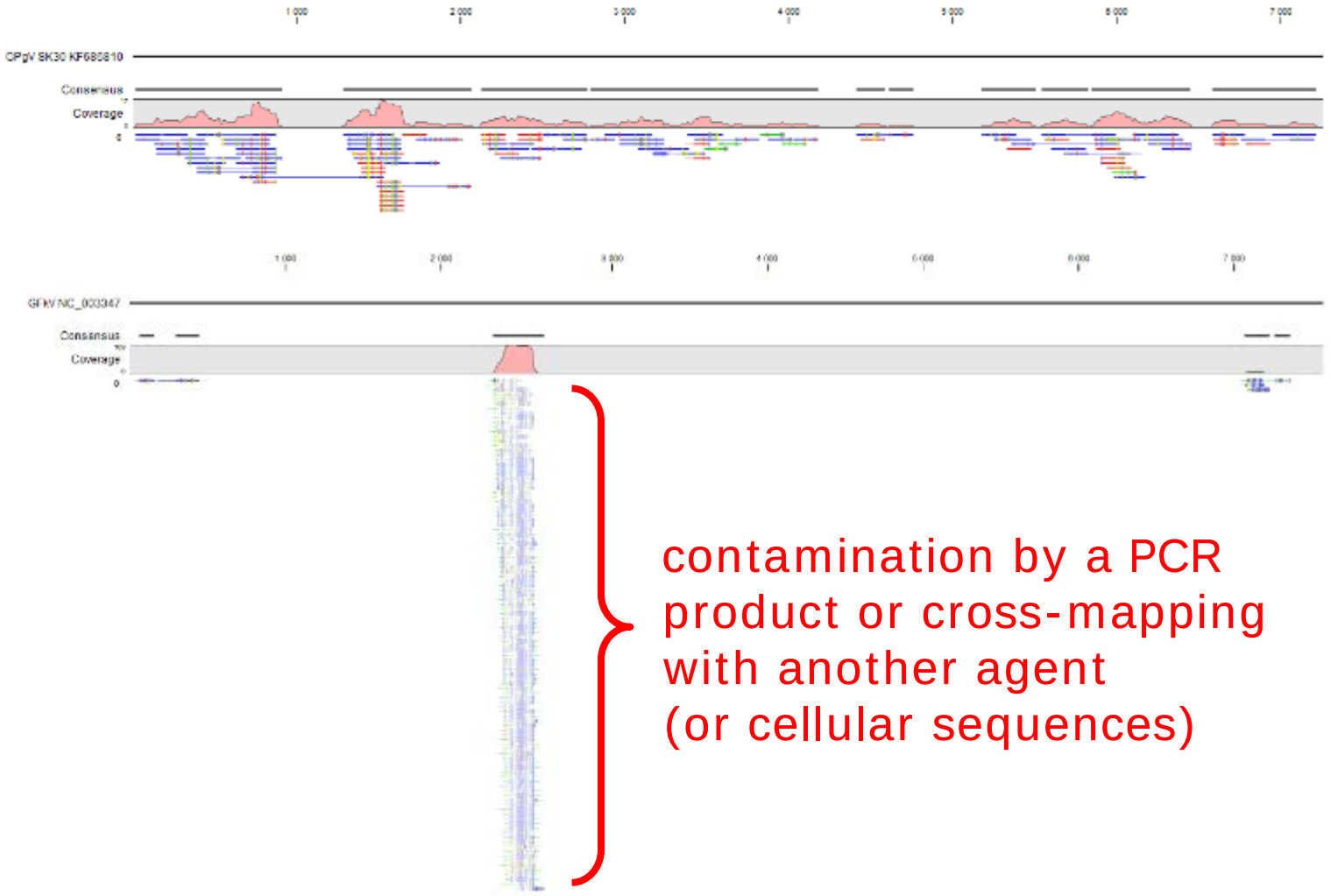
LChV1, ACLSV, PDV, new *Tepovirus* (*Betaflexiviriae*)

All mapped reads are not equal....

- **Grapevine sample, RNASeq, mapping analysis**
  - 145 Reads *Grapevine Pinot gris virus* (GPgV)
  - 116 reads *Grapevine fleck virus* (GFkV)

GPgV

GFkV



contamination by a PCR product or cross-mapping with another agent (or cellular sequences)

Mapping stringency matters...

Grapevine, total RNASeq, mapping of reads against reference database

	Reads mapped using various stringencies			
	100%	90%	80%	70%
GFkV	0	5	17	401
GRVFV	0	203	4301	4774

Mapping stringency may ultimately need to be fine tuned for each virus, taking into account its variability and that of related agents

But

* Very difficult and time consuming to use different stringency parameters for each virus >> use compromise
* Our knowledge of viral variability is incomplete.... (new divergent strains regularly detected....)
* In some cases, it may not be possible to select an optimal stringency (interspecific recombinant viruses...)

Accuracy

- **False negatives ? cf sensitivity & specificity**
  - Performance of bioinformatic pipeline
  - **COST proficiency test**: only **2/12** pathogens detected by all pipelines at highest sequencing depth
  - Novel viruses too divergent to recognize by a Blast-based approach ? Additionnally use motive searches (HmmScan…)
- **False positives ? cf specificity plus other issues**
  - Need for expertise when looking at pipeline results
  - For DNA viruses, integrated or episomal virus ??? *(in particular for Caulimoviridae members in RNAseq)*
  - Contamination (diagnostic lab or sequencing platform). NGS at least as susceptible to contaminations as PCR (one or more PCR step(s) in most NGS protocols)

Example of contamination

Ribo-depleted RNASeq on germinating radish seedling

	Reads/virus/10^6 reads				
	New Bunyaviridae	New Cryptic virus	RsCV2	PhMCV	PvEV1
Radish #1	25828	59213	256	1466	1515
Radish #2	27164	120436	1280	685	1799
PCR on seedlings	-	+	+	-	-

Novel *Bunyaviridae*

Near complete genome assembled

Origin of contamination ???

Sequencing platform ?

Physotegia chlorotic mottle virus

Complete genome assembled

Origin of contamination ???

Sequencing platform ?

Phaseolus endornavirus 1

Complete genome

Likely lab contamination, frequent bean samples

(Data M. Barret, IRHS Angers, France)

Conclusions

- NGS technologies have already drastically changed virus discovery and etiology
- They have the potential to drastically change the field of viral indexing/diagnostics, providing **faster, cheaper and more complete results**, with many applications
  - Certification/quality control
  - Quarantine
- There are pitfalls and challenges, NGS is more complicated than PCR and some expertise is needed
- Similar to the situation with PCR in its early years, much work still needed for full adoption in diagnostics
  - **Comparative performance** with existing assays *(sensitivity, repeatability…)*
  - Validated protocols, including **data analysis** & **detection thresholds** *(Benchmarking, Proficiency tests, Ringtests…)*
  - **Quality management** systems, Standards…

Many thanks to…

- **Bordeaux team**
  - A. Marais
  - C. Faure
  - L. Svanella
  - S. Theil

- **All COST DIVAS members and in particular**
  - S. Massart
  - T. Wetzel
  - A. Olmos

- Many other colleagues and collaborators with whom I exchanged ideas…


GRAZIE MILLE !
THANK YOU !