Introduction of the webinar series

The use and validation of High Throughput sequencing (HTS) tests for diagnostics of plant pests.

Webinar 1	What is High Throughput Sequencing (HTS)?	Friday 30th April, 2 pm-3:30 pm
Webinar 2	How to prepare your laboratory to conduct HTS tests?	Monday 3rd May, 2 pm-3 pm
Webinar 3	How to develop, validate and routinely use HTS for diagnostic purpose?	Tuesday 4th May, 2 pm-3:30 pm
Practical training activity	How to apply the guidelines to your laboratory?	Wednesday 5th May, 2 pm to 4:30 pm. Friday 7th May, 2 pm to 4:30 pm.

Valitest

LIÈGE université Gembloux Agro-Bio Tech

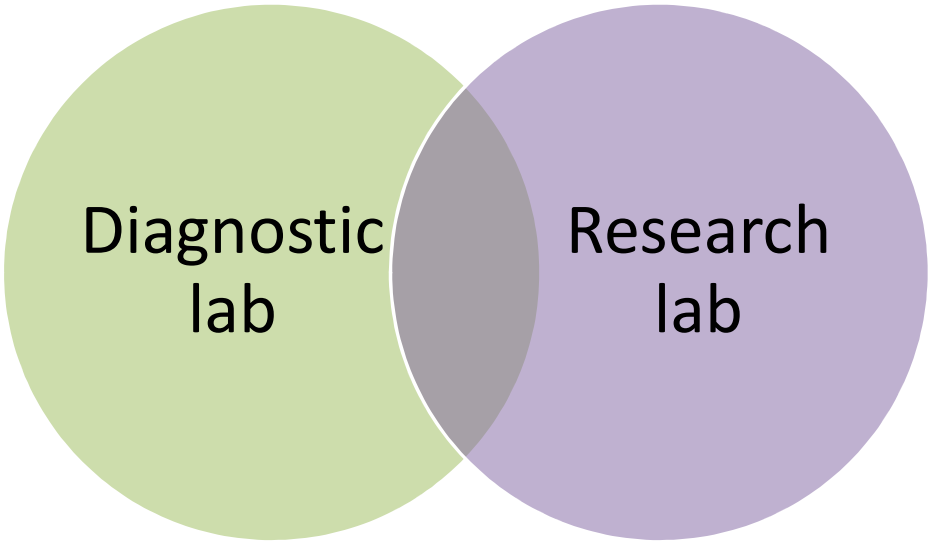
Preparing the laboratory for the use of HTS: General and technical requirements

Sebastien Massart

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 773139

Valitest

1. Introduction



1. Introduction : diagnostic use of HTS

Bulletin OEPP/EPPO Bulletin (2018) **0** (0), 1–6 ISSN 0250-8052. DOI: 10.1111/epp.12472

High-throughput sequencing technologies for plant pest diagnosis: challenges and opportunities

A. Olmos[1], N. Boonham[2], T. Candresse[3], P. Gentit[4], B. Giovani[5], D. Kutnjak[6], L. Liefting[7], H.J. Maree[8], A. Minafra[9], A. Moreira[10], M.K. Nakhla[11], F. Petter[5], M. Ravnikar[6], B. Rodoni[12], J.W. Roenhorst[13], M. Rott[14], A.B. Ruiz-García[14], J. Santala[15], G. Stancanelli[16], R. van der Vlugt[17], C. Varveri[18], M. Westenberg[13], T. Wetzel[19], H. Ziebell[20] and S. Massart[21]

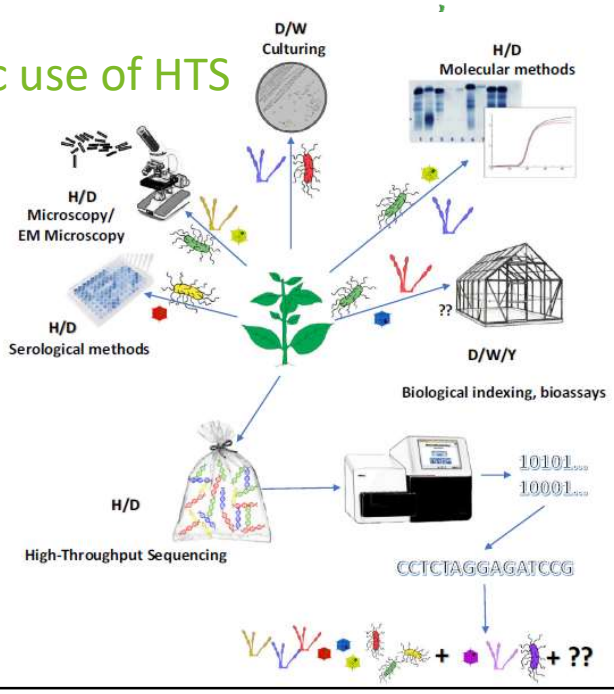
Eur J Plant Pathol
https://doi.org/10.1007/s10658-018-1570-0

CrossMark

SI: PLANT PATHOLOGY FOR INNOVATIVE AGROECOLOGY

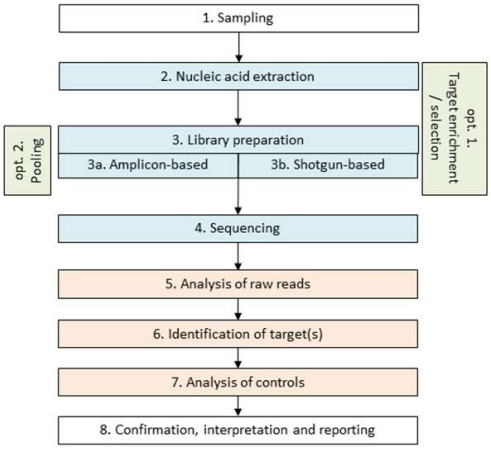
The impact of high throughput sequencing on plant health diagnostics

Ian P. Adams · Adrian Fox · Neil Boonham · Sébastien Massart · Kris De Jonghe



BL1

Overview of the HTS process in plant health diagnostic



1. Introduction: focus on requirements

- General
- Technical

www.wooclap.com/OCIPBU

Slide 5

BL1 cette figure est 'boring', ce serait mieux avec des images, du coup je propose celles du papier de Piper (slide suivante) mais je ne sais pas si tu as le droit de t'en servir

Bénédicte Lebas; 28-04-21

LIÈGE université
Gembloux
Agro-Bio Tech

2. Focus on the general requirements:

« The frame of the house »

2.1 Infrastructure

- Laboratory & Informatics
  - *The laboratory should have, or have access to, appropriate facilities and information technology (IT) infrastructures to perform HTS tests.*
- One-way process in the lab (PCR-like)

2.1 Infrastructure: lab

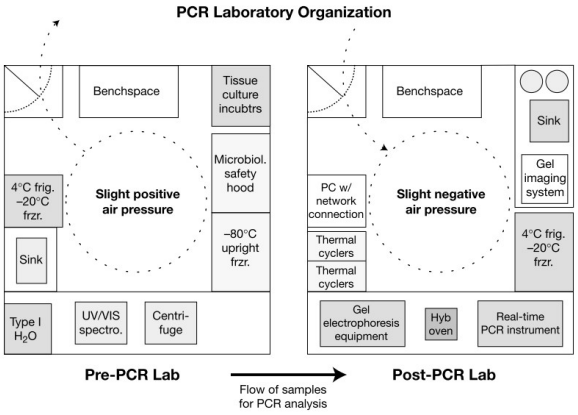


FIGURE 2. Organization of a PCR laboratory with separate pre- and post-PCR rooms.

Mifflin. Setting up a PCR laboratory. www.biosupplynet.com

+ room for reagent mix

+ Several PCR (amplicon)

+ (sequencing machine)

EPPO standard PM 7/98 (2019)

2.1 Infrastructure: lab

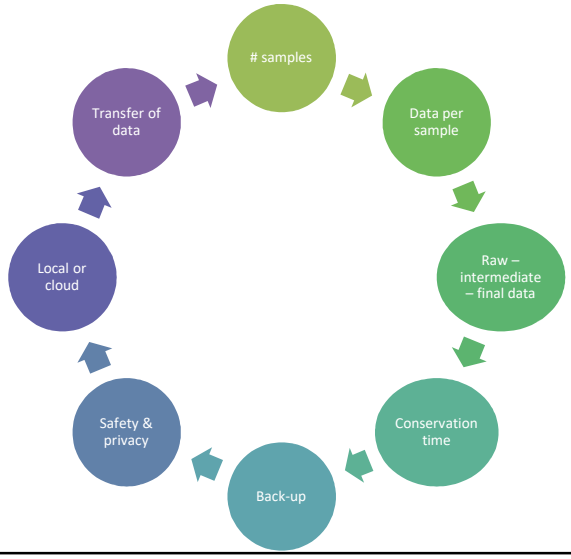
- See EPPO standard 7/98 (2019) recommendations for molecular biology

2.1 Infrastructure: IT

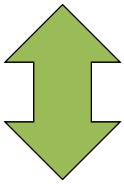
- New aspect !
- Gigabite era
- What are the key elements to take into account for IT infrastructure ?

www.wooclap.com/OCIPBU

2.1 Infrastructure: IT



Storage (hard drive)



Computing power (cpu)

2.2 Personnel

- For scientist/manager & technicians
  - *The laboratory should have personnel proven to be competent for the overall HTS process including laboratory and bioinformatic components and for the biological interpretation of the data.*
- For the laboratory
- For the data management (IT)
- For the analysis of the data (bioIT)
- For the interpretation of the results (scientist)

2.3 Quality Management system

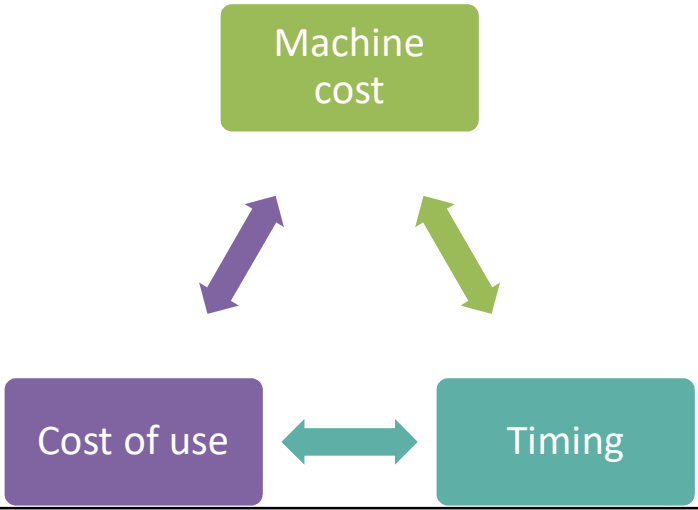
- Laboratory & IT
  - *The laboratory should have a quality management system in place to ensure its consistent operation in performing HTS tests and traceability throughout the process.*
  - *ISO 17025 accreditation in EU*

2.4 Outsourcing: sequencing

- Illumina examples:
  - NovaSeq 6000 : ~985,000 € & ~3 €/M reads
  - Miseq : ~125,000 € & ~100 €/M reads

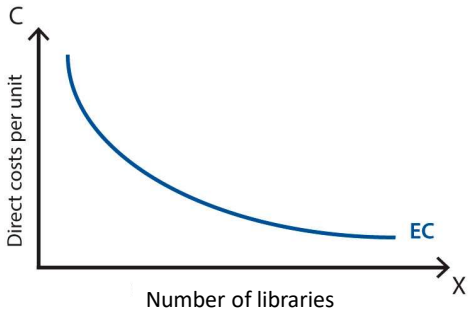
- Economical optimum: outsourcing to facility

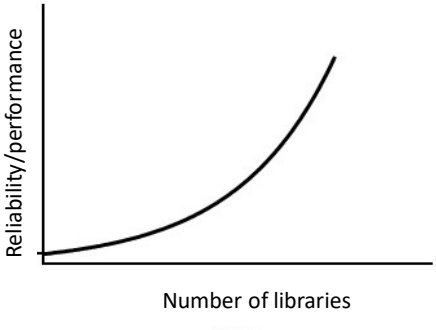
2.4 Outsourcing: sequencing



2.4 Outsourcing: library prep ?

- Experience curve





2.4 Outsourcing

- If outsourcing:
  - *The laboratory should check the quality performance of the outsourced services at appropriate intervals*
- Audit or official accreditation (sequencing service provider ISO 17025 accredited)

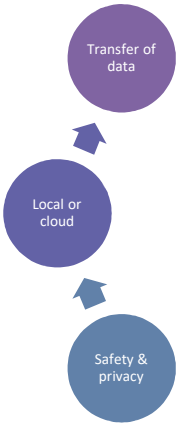
2.5 Managing modifications

- Laboratory and bioIT protocols =

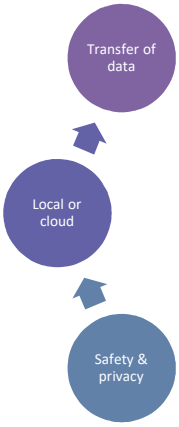
- Recommendation:
  - *The laboratory should have a procedure for monitoring, implementing, and documenting modifications of reagents, kits, sequencing chemistries, instruments and bioinformatic*

2.6 Data management

- Recommendation:
  - *The laboratory should have access to a network allowing a safe, easy and complete transfer of data generated during each HTS test with an appropriate backup and storage system.*



2.6 Data management



- Reference databases:
  - *The laboratory should use appropriate sequence databases to analyse the sequence data generated by the HTS test*
  - ***Completeness* vs. *curation***

LIÈGE université
Gembloux
Agro-Bio Tech

3. Focus on the technical requirements:

« The content of the house »

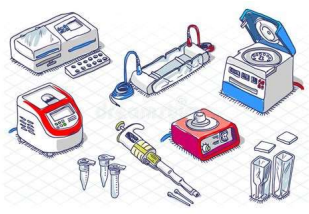
Laboratory

3.1 Scope of the test

- Scope of HTS tests
  - *The laboratory should define the scope of the HTS test*
- Scope = matrix and targets

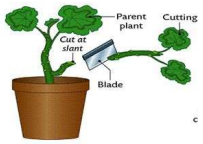
3.2 Laboratory component

- From Sampling to sequencing
  - Each step should be selected, developed and optimized and validated for its intended use
  - After the validation of an HTS test, its performance should be monitored using appropriate controls during its routine use



3.2 Laboratory component - Sampling

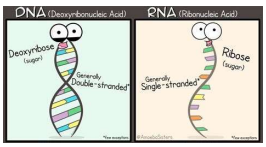
- Sampling
  - *The sampling protocol should be appropriate to the matrix and pest(s) targeted by the HTS test.*

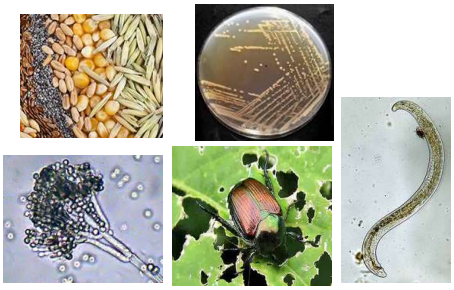


- Sample handling
  - *Sample handling should ensure sample integrity and suitability for the HTS test*

3.2 Laboratory component – Nucleic acids

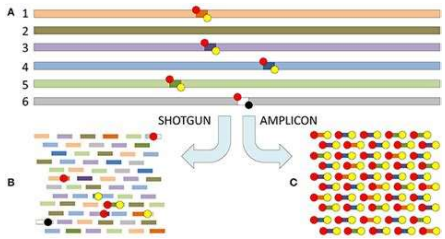
- Sampling
  - *The nucleic acid extraction protocol should deliver nucleic acids of appropriate quality and quantity for the HTS test.*





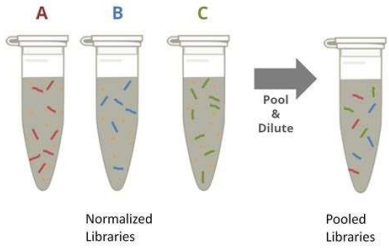
3.2 Laboratory component – Library prep

- Recommendation
  - *The library preparation protocol should be suitable for the sequencing platform.*



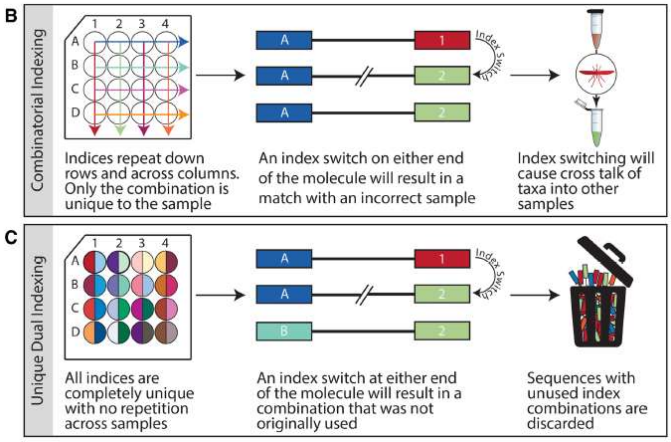
3.2 Laboratory component – Library prep

- Pooling of libraries
  - *The pooling level of libraries should be adapted to the intended use of the HTS test and its required analytical sensitivity.*



3.2 Laboratory component – Library prep

- Pooling of libraries



Piper et al. 2019. GigaScience, 8, 1-22

3.2 Laboratory component – Sequencing

- Recommendation
  - *The sequencing platform and method should be appropriate for the intended use of the HTS test.*
- Comparison :

	Illumina	Ion torrent	PacBio	Minion
Detection method	Fluorescence	Electric signal	Fluorescence	Electric signal
Nucleotides	Mix	One by one	Mix	n.a.
Enzyme	Reagent	Reagent	Fixed	Fixed
Amplification	Bridge	Beads	No	No
Lenght	Max 300 nt	Max 300 nt	Up to 10 kb	Up to 10 kb
Fixation of DNA	Oligo on cell	Oligo on beads	Enzyme	Enzyme + nanopore

3.2 Laboratory component – Sequencing

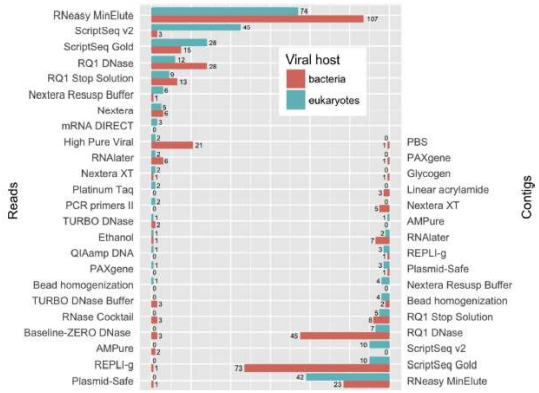
- expected number of samples received per batch and reads number per sample,
- required test turn-around time (e.g. urgent testing for imported perishable material),
- total number of generated reads per sequencing run,
- compatible with the expected number of samples per batch~~?~~
- read length and type (e.g. single, paired, mate-pair): short single reads are appropriate for sRNA sequencing whereas amplicon sequencing might need longer reads,
- error rate and type of error which vary between sequencing platforms and between runs,
- impact on the downstream bioinformatic analyses (depending on the number of sequences, their length, their quality and accuracy),
- availability of bioinformatic support, laboratory resources and technical expertise and manufacturer level of technical support,
- expenses involved in the operation of a sequencing machine: purchase and maintenance (Rehm *et al.*, 2013).

3.2 Laboratory component – Contamination

- Recommendation:
  - *The laboratory should prevent contaminations as they can critically impact the results of HTS tests.*
- Personal observation:
  - « ppm » contamination
  - From qualitative to quantitative approach

3.2 Laboratory component – Impact

- Impact of viral/bacterial contaminant in HTS datasets



Asplund et al. 2019, Clinical Microbiology and Infection, 25(10), 277-1285

LIÈGE université
Gembloux
Agro-Bio Tech

3. Focus on the technical requirements:

« The content of the house »

Bioinformatics

3.1 Does bioinformatics matter ?

Evaluate and compare the performance of bioinformatic pipelines for the analysis of high-throughput siRNA sequencing data

EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

DIVAS COST Action FA 1407

3.1 Does bioinformatics matter ?

Species	Potato		Apple	Grapevine									
Sequencing technology	Illumina												
Sequencing depth	23M		13M	12M									
Plant	Potato		Apple	Grapevine rootstock Kober 125 aa, Vitis berlandieri x Vitis riparia									
Viruses or viroids detected (one column per virus)	PVX	New Nepovirus	ASGV (2x genetic variants)	Grapevine leafroll-associated virus 1	Grapevine virus A	Grapevine virus B	Grapevine rupestris stem pitting-associated virus	Grapevine rupestris vein feathering virus	Grapevine Syrah virus 1	Grapevine red globe virus	Hop stunt viroid	Grapevine yellow speckle viroid 1
Percentage of reads for each virus/viroid (*)	~23%	~5%	0.12%	5.7%	0.6%	1.2%	0.4%	0.3%			0,5%	0,14%
Confirmation(s) technique(s) for each virus/viroid	RT-PCR & Sanger sequencing	RT-PCR & Sanger sequencing	RT-PCR & Sanger sequencing	ELISA, RT-PCR	ELISA, RT-PCR	ELISA, RT-PCR	RT-PCR	not yet verified = not tested	RT-PCR	negative in RT-PCR	RT-PCR	RT-PCR

Difficulties:

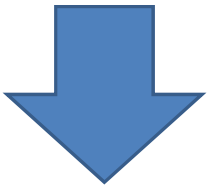
1. **New unknown virus**
2. **Complex mix of 9 viruses/viroids**
3. **Very low abundance virus/viroid**
4. **Close virus species with poor genome knowledge**

3.1 Does bioinformatics matter ?

- sRNA datasets (lenght between 21 and 24 nt)
- Worst case scenario by rarefaction at 3 sequencing depths:
  - 50,000 reads – 3 files
  - 250,000 reads– 4 files (2 grapevine replicates)
  - 2,500,000 reads– 3 files
- 10 fastq files available on a server in double blind
- 21 participants

3.1 Does bioinformatics matter ?

Participants free to apply their own bioinformatics strategy to identify the viruses in the 10 datasets (de novo assembly + annotation)



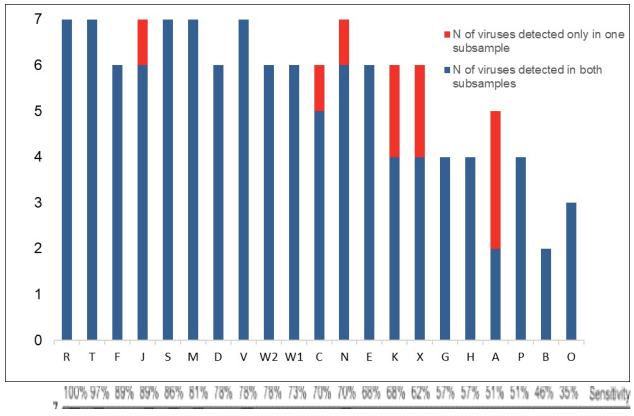
Interpret the data in a diagnostics setting

3.1 Does bioinformatics matter ?

	Sensitivity			
labID	50,000	250,000	2,500,000	Average
A	10%	53%	90%	51%
B	30%	35%	80%	46%
C	60%	71%	80%	70%
D	50%	82%	100%	78%
E	30%	82%	80%	68%
F	80%	88%	100%	89%
G	20%	53%	100%	57%
H	30%	65%	70%	57%
J	70%	94%	100%	89%
K	40%	71%	90%	68%
M	50%	94%	90%	81%
N	30%	82%	90%	70%
O	20%	41%	40%	35%
P	20%	59%	70%	51%
R	100%	100%	100%	100%
S	50%	100%	100%	86%
T	90%	100%	100%	97%
V	60%	88%	80%	78%
W1	40%	82%	90%	73%
W2	60%	82%	90%	78%
X	30%	71%	80%	62%
AVERAGE	46%	75%	86%	70%

- Analytical sensitivity
  - 70%
  - From 35% to 100 %
    - FDR !
  - Number of reads
  - 1/3 : 100% at 2.5M

3.1 Does bioinformatics matter ?



- Reproducibility
  - 92 %
  - 100% for 15 strategies
  - Correlation with analytical sensitivity

3.1 Does bioinformatics matter ?

- YES !

Phytopathology • 2019 • 109:488-497 • https://doi.org/10.1094/PHYTO-02-18-0067-R

Virology

e-Xtra*

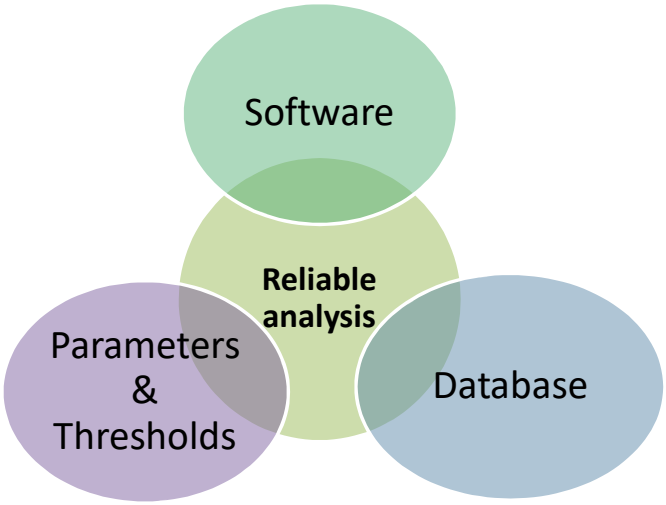
Virus Detection by High-Throughput Sequencing of Small RNAs: Large-Scale Performance Testing of Sequence Analysis Strategies

Sebastien Massart,† Michela Chiumenti, Kris De Jonghe, Rachel Glover, Annelies Haegeman, Igor Koloniuk, Petr Komínek, Jan Kreuze, Denis Kutnjak, Leonidas Lotos, François Maclot, Varvara Maliogka, Hans J. Maree, Thibaut Olivier, Antonio Olmos, Mikhail M. Pooggin, Jean-Sébastien Reynard, Ana B. Ruiz-García, Dana Safarova, Pierre H. H. Schneeberger, Noa Sela, Silvia Turco, Eeva J. Vainio, Eva Varallyay, Eric Verdin, Marcel Westenberg, Yves Brostaux, and Thierry Candresse

3.2 Bioinformatics component ?

- Key element of the HTS test as it can generate false positive and/or false negative results
- Consists of a combination of software used to analyse the raw sequencing data

3.2 Bioinformatics component ?



3.2 Bioinformatics component ?

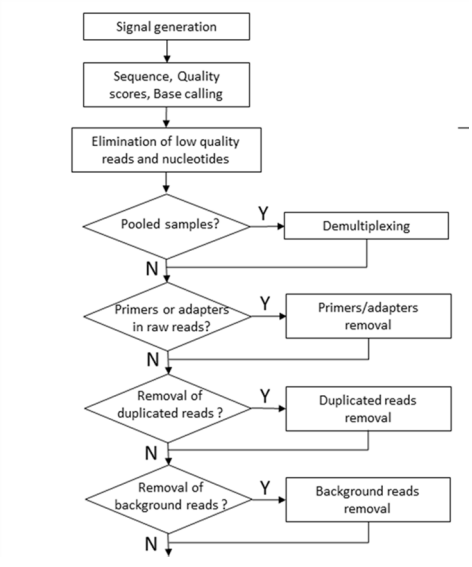
- Divided in three steps with several sub-steps :

1. Analysis of raw reads
2. Identification of targets / pests
3. Analysis of controls

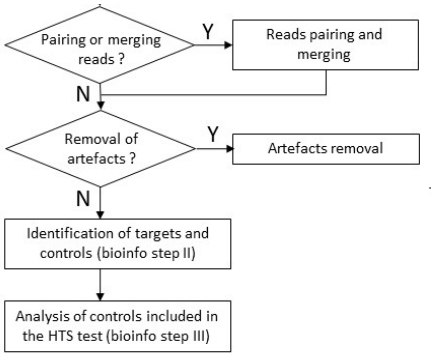
3.3 Analysis of raw reads

- Recommendation:
  - *The laboratory should eliminate low quality sequences and assign unambiguously the sequences to each sample. Optionally, redundant or background sequences can be eliminated.*
- A flowchart has been drafted to guide this step

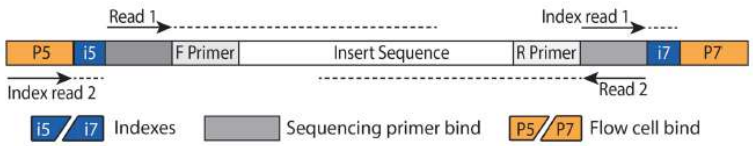
3.3 Analysis of raw reads

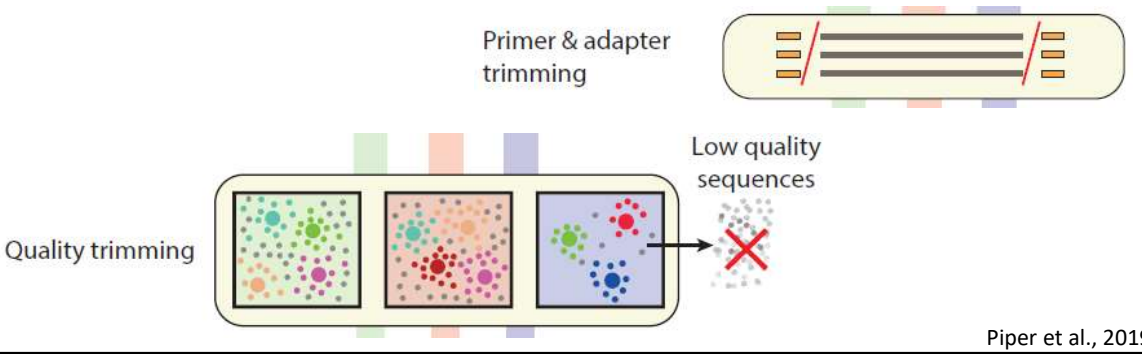


3.3 Analysis of raw reads



3.3 Analysis of raw reads





Piper et al., 2019

3.3 Analysis of raw reads

- Misassignment during demultiplexing: barcode switching / errors

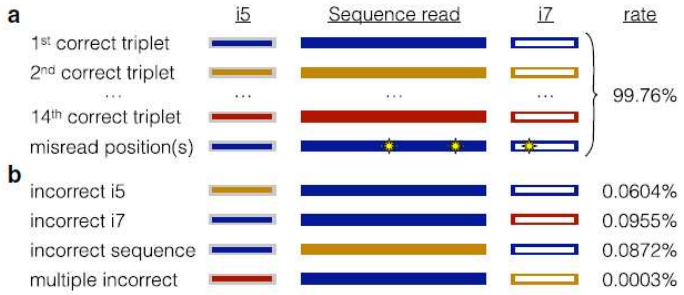


Fig. 1 Rates of different misassignment errors on the Illumina platform. **a** Unique index and read sequences that were well separated in sequence space (colored rectangles) were used to form 14 distinct samples and multiplexed in the same Illumina sequencing run. Misread bases (yellow stars) make up the most common error type, but are still attributable to their correct triplet. **b** Misassigned reads appear as unexpected triplets, and can be categorized as either index misassignments (0.16 % total) or sequence misassignments (0.09 %)

Wright and Vetsigian BMC Genomics (2016) 17:876

3.3 Analysis of raw reads

- Trade off between removing misassigned and preserving correct reads during quality filtering

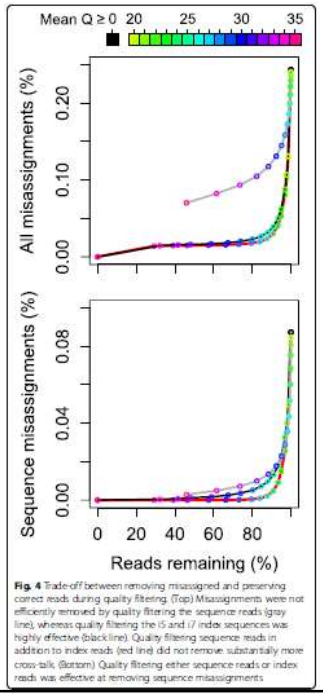


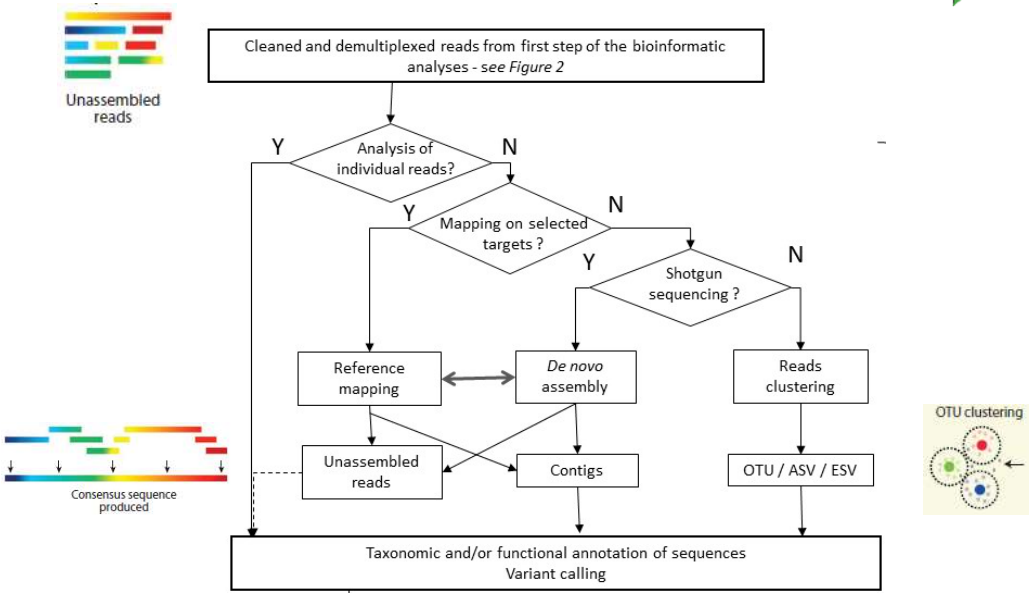
Fig. 4 Trade-off between removing misassigned and preserving correct reads during quality filtering. (Top) Misassignments were not efficiently removed by quality filtering the sequence reads (gray line), whereas quality filtering the i5 and i7 index sequences was highly effective (black line). Quality filtering sequence reads in addition to index reads (red line) did not remove substantially more cross-talk. (Bottom) Quality filtering either sequence reads or index reads was effective at removing sequence misassignments

Wright and Vetsigian BMC Genomics (2016) 17:876

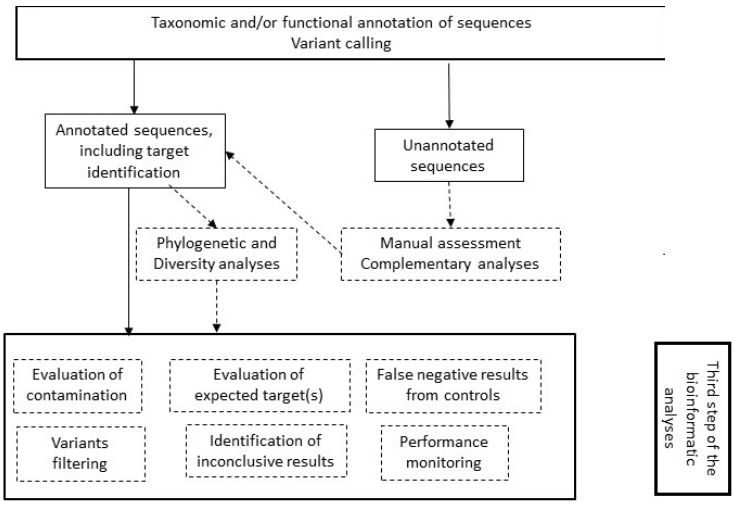
3.4 Identification of targets

- Recommendation :
  - *The laboratory should identify target(s) of the analysed samples and controls from the cleaned reads.*

3.4 Identification of targets



3.4 Identification of targets



3.5 Analysis of controls

- Recommendation:
  - *The laboratory should verify that all the controls used in the HTS test performed as expected.*
  - Evaluation of expected target(s)
  - False negative results from controls
  - Evaluation of contamination

LIÈGE université
Gembloux
Agro-Bio Tech

4. Conclusion

4. Conclusion

- HTS technologies open new possibilities and opportunities in routine diagnostics
- Not an easy change (serology -> molecular biology)
- Similar general and technical requirements to any molecular methods
  - Laboratory complexity (steps)
  - Informatic and bioinformatics components

LIÈGE université
Gembloux
Agro-Bio Tech

Valitest

Thank you for your attention
& your participation to the pools !

Sebastien Massart
sebastien.massart@uliege.be

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 773139

@Be_Phytopath