Metagenòmica Clínica per al diagnòstic de malalties infeccioses

F. Xavier López-Labrador

Laboratori de Virologia, FISABIO Salut Pública, Generalitat Valenciana

Secc. Depart. Microbiologia, Facultat de Medicina, Universitat de Valencia

CIBERESP, Instituto de Salud Carlos III

F.Xavier.Lopez@uv.es

XXXI JORNADES SOCIETAT CATALANA DE MALATIES INFECCIOSES I MICROBIOLOGÍA CLÍNICA.
Montbrió del Camp *14-15 OCT. 2022*

"Omics" Era: Genomics

- Microbiome analyses
- Environmental / animal / clinical metagenomics
- Pathogen discovery
- Public health disease surveillance (outbreak investigation)
- Massively multiplex testing applications (detection of adventitious agents in food/water; blood screening)
- Clinical infectious disease diagnosis by metagenomic next-generation sequencing (mNGS)
  - *metagenomic sequencing, unbiased sequencing, agnostic sequencing, non-targeted sequencing, shotgun sequencing*

2002

Genomic analysis of uncultured marine viral communities

Mya Breitbart*, Peter Salamon†, Bjarne Andresen†‡, Joseph M. Mahaffy†, Anca M. Segall*, David Mead§, Farooq Azam¶, and Forest Rohwer*||

*Department of Biology, San Diego State University, San Diego, CA 92182-4614; †Department of Mathematical Sciences, San Diego State University, San Diego, CA 92182-7720; ‡Orsted Laboratory, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark; §Lucigen, Middleton, WI 53562; and ¶Marine Biology Division, Scripps Institution of Oceanography, La Jolla, CA 92093

Communicated by Allan Campbell, Stanford University, Stanford, CA, August 14, 2002 (received for review February 22, 2002)

Viruses are the most common biological entities in the oceans by an order of magnitude. However, very little is known about their diversity. Here we report a genomic analysis of two uncultured marine viral communities. Over 65% of the sequences were not significantly similar to previously reported sequences, suggesting that much of the diversity is previously uncharacterized. The most common significant hits among the known sequences were to viruses. The viral hits included sequences from all of the major families of dsDNA tailed phages, as well as some algal viruses. Several independent mathematical models based on the observed number of contigs predicted that the most abundant viral genome

by pulse field gel electrophoresis (11); however, this method will not recover all viruses (e.g., large eukaryotic viruses and ssRNA phages). After CsCl purification, the viruses were lysed by using a formamide extraction, and the DNA was recovered by an isopropanol precipitation and a cetyltrimethylammonium bromide (CTAB) extraction (12).

Construction of the Shotgun Library. The amount of viral DNA in an environmental sample is very low (~10 μg/100 liters). Viral genomes often contain modified nucleotides that cannot be directly cloned into *Escherichia coli*. Additionalb, because viral

The NEW ENGLAND JOURNAL of MEDICINE

2019

ORIGINAL ARTICLE

Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis

M.R. Wilson, H.A. Sample, K.C. Zorn, S. Arevalo, G. Yu, J. Neuhaus, S. Federman, D. Stryke, B. Briggs, C. Langelier, A. Berger, V. Douglas, S.A. Josephson, F.C. Chow, B.D. Fulton, J.L. DeRisi, J.M. Gelfand, S.N. Naccache, J. Bender, J. Dien Bard, J. Murkey, M. Carlson, P.M. Vespa, T. Vijayan, P.R. Allyn, S. Campeau, R.M. Humphries, J.D. Klausner, C.D. Ganzon, F. Memar, N.A. Ocampo, L.L. Zimmermann, S.H. Cohen, C.R. Polage, R.L. DeBiasi, B. Haller, R. Dallas, G. Maron, R. Hayden, K. Messacar, S.R. Dominguez, S. Miller, and C.Y. Chiu

Clinical metagenomic sequencing (mNGS)

- Metagenomics is sequencing all DNA and RNA in a sample
- Don't need to know what to look for
- Any virus, bacteria, fungi, parasite in one shot
- Can detect unusual, unexpected, divergent pathogens
- mNGS is increasingly being used in virology laboratories for difficult to diagnose cases

Clinical metagenomic sequencing (mNGS)

- Meningitis / Encephalitis: *40-60% unknown cause*
- Pneumonia: *15 – 25% unknown cause*
- Fever / Sepsis *~20% unknown cause*
- Failure to obtain a timely diagnosis leads to delayed / inappropriate therapy, increased mortality, and excess healthcare costs
- Current main clinical application of mNGS is meningitis / encephalitis, but considered useful in a growing number of other clinical syndromes (pneumonia, sepsis, etc.)

Diagnostic approach

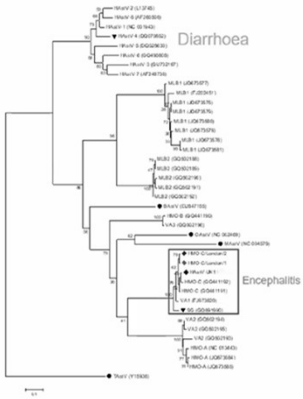
Single target	Syndromic panels	mNGS
HSV-1, 2; VZV; ENV, HHV-6… (IS): CMV, ADV, EBV, JCV, Flu, Mumps/Measles	Gastro panel Respiratory panel Parvo/Rubella/Paraecho	**All pathogens**

25+ targets, Sufficient specimen? One sample for all

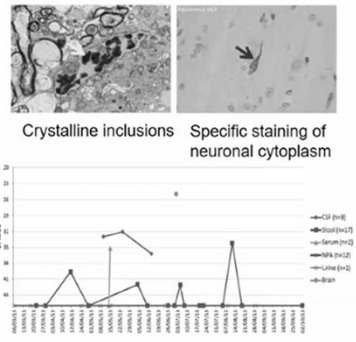
Astrovirus Encephalitis revealed by mNGS

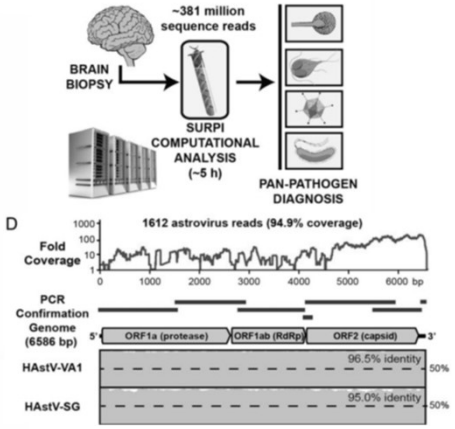
HAstV-VA1 now a recognized cause of encephalitis in immunocompromised

"Encephalitic" Astrovirus Clade in Humans



Brown, CID 2015



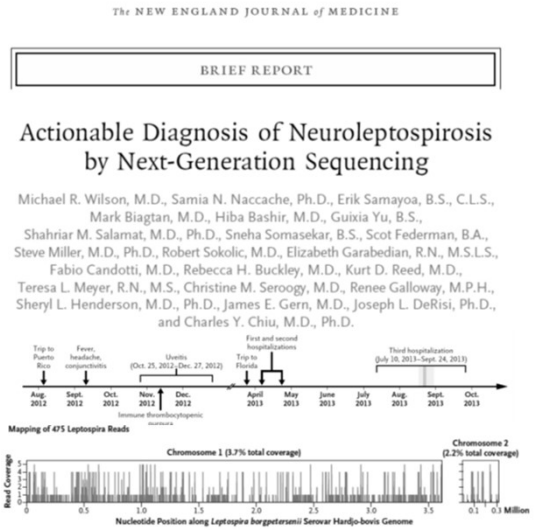


Naccache, CID 2015

Neuroleptospirosis revealed by mNGS

- 14 y.o. Male, SCID
- Fever, headache
- 3 hospitalizations in 4 months
- Progressed to hydrocephalus, status epilepticus
- 44 days in ICU
- Brain biospsy, induced coma
- >100 tests (inconclusive)
- **mNGS in CSF**:

Leptospira santarosai
Leptospira borgpetersenii
unclassified
Leptospira interrogans
Propionibacterium acnes



ENNGS from 2018 to date

ESCV
European Society for Clinical Virology

ENNGS
ESCV NETWORK ON NEXT-GENERATION SEQUENCING

- **Workshop on viral metagenomics**
  - 2018: Leiden, NL
  - 2019: Valencia, ES
  - 2020: cancelled due to COVID pandemic
  - 2021: cancelled due to COVID pandemic
  - 2022: Antalya, Turkey
- **Recommendation paper Part I on wet lab procedures:**
  López-Labrador et al. J Clin Virol 2021
- **Recommendation paper Part II on dry lab procedures:**
  de Vries et al. J Clin Virol 2021
- **Benchmark of metagenomic pipelines for viral pathogen detection:**
  de Vries et al. J Clin Virol 2021

Journal of Clinical Virology 134 (2021) 104691

Contents lists available at ScienceDirect

Journal of Clinical Virology

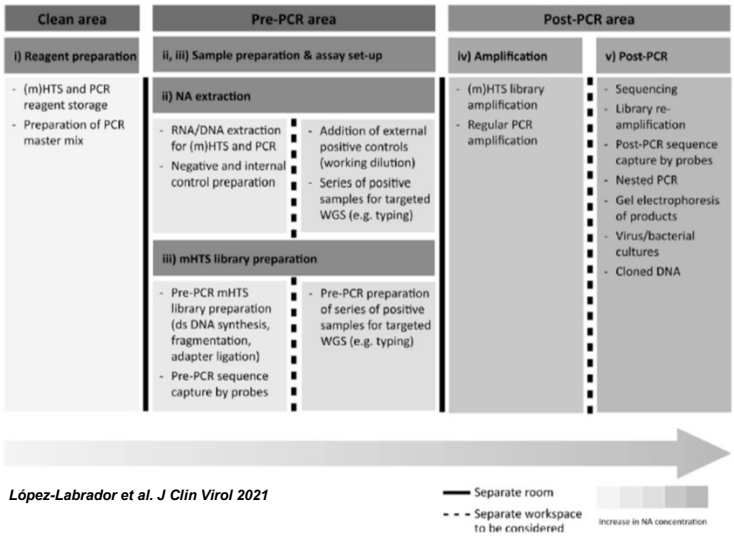
journal homepage: www.elsevier.com/locate/jcv

Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: Wet lab procedure

F. Xavier López-Labrador [a,b], Julianne R. Brown [c], Nicole Fischer [d], Heli Harvala [e], Sander Van Boheemen [f], Ondrej Cinek [g], Arzu Sayiner [h], Tina Vasehus Madsen [i], Eeva Auvinen [j], Verena Kufner [k], Michael Huber [k], Christophe Rodriguez [l], Marcel Jonges [m,n], Mario Hönemann [o], Petri Susi [p], Hugo Sousa [q,r,s,t], Paul E. Klapper [u], Alba Pérez-Cataluña [v], Marta Hernandez [w], Richard Molenkamp [f], Lia van der Hoek [m,n], Rob Schuurman [x], Natacha Couto [y,z], Karoline Leuzinger [A,B], Peter Simmonds [C], Martin Beer [D], Dirk Höper [D], Sergio Kamminga [E], Mariet C.W. Feltkamp [E], Jesús Rodríguez-Díaz [F], Els Keyaerts [G], Xiaohui Chen Nielsen [i], Elisabeth Puchhammer-Stöckl [H], Aloys C.M. Kroes [E], Javier Buesa [F], Judy Breuer [c], Eric C. J. Claas [E], Jutte J.C. de Vries [E,*], on behalf of the ESCV Network on Next-Generation Sequencing

Facility / floor plan: requirements for diagnostics

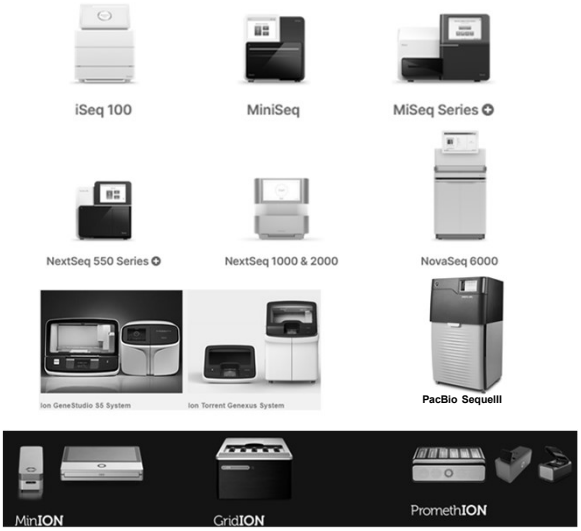
1 Physical separation of reagent preparation, pre-amplification and post-amplification library preparation.
2 Dedicated materials and reagents for each process (sample processing, library preparation, post-library preparation).
3 Physical separation of metagenomic library preparation from sample preparation of series of positive samples (e.g. for typing), e.g. by using a dedicated biosafety cabinet (BSC) with restricted use for metagenomic workflows.
4 Extensive cleaning of materials and surfaces with 10% sodium hypochlorite and/or ammonium compound before and after processing, more frequently than regularly performed for molecular assays.



López-Labrador et al. J Clin Virol 2021

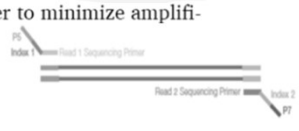
Sequencing platform: which one?

5 Choice depending on the application and intended use (metagenomics, whole genome sequencing, fieldwork).
6 Restrict low output sequencers use for a limited number of specimens (due to their lower throughput and multiplexing and deep sequencing capacity).
7 Consider the number of samples per run in relation with batch-wise sequencing and consequences for turn-around-time. Outsourcing of parts of the process may be carefully considered with attention to quality, safety, transparency and flexibility to desired adaptations of the protocol.



Assay design and development

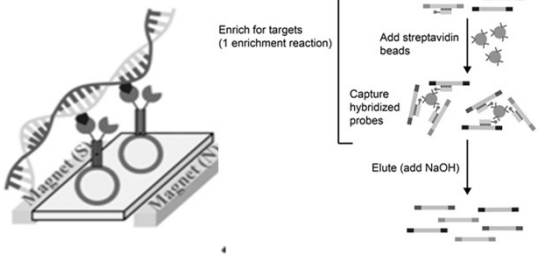
8 DNA and RNA can be co-extracted or isolated separately, with impact on the mHTS results (sensitivity, coverage), and separate protocols should be validated individually.
9 Avoid the use of high concentrations of carrier RNA during extraction for RNA mHTS.
10 Advantages of target enrichment should be weighed against the potential bias introduced by the specific protocol.
11 SISPA and MDA should not be used when performing viral metagenomics aiming at quantification of viral species, since this may result in over- and underrepresentation of the true proportions for certain viruses.
12 The minimum number of post-ligation amplification cycles should be used, in order to minimize amplification bias.



http://www.nature.com/scitable/topicpage/toll-like-receptors-sensors-that-detect-infection-14396559

DNAseq improves detection of DNA viruses (and bacteria / fungus / parasite)

Validate enrichment



Assay design and development

13 The library size distribution should be checked for the expected fragment size, to discard degraded libraries (excess short fragments) or incomplete fragmentation (excess long fragments). Accurate library quantitation ensures adequate library pooling in the sequencing run.

14 A no-template control that will undergo all steps from sample extraction to sequencing should be used in every individual sequencing run.

15 More upfront negative controls are recommended to identify sources of potential contamination, such as a library preparation buffer and a pathogen-negative sequence controls (e.g. phage lambda prepared with different reagents).

16 To control for the success of NA extraction, preparation and sequencing, clinical samples should be spiked with encapsidated RNA or DNA viruses that do not infect humans (vertebrates), e.g. bacteriophages.

Reagents contain multitudes of (bacterial) contaminants
Normalize to eliminate background:
(Ratio of Patient / NTC)

Low / no reads in spiked-in controls =
Low reads (< 5M)
Inadequate DNA data / high background
Inadequate RNA data / degradation?

ESCV European Society for Clinical Virology — ENNGS ESCV NETWORK ON NEXT-GENERATION SEQUENCING

GENERALITAT VALENCIANA · Fundació Fisabio · CENTRE SUPERIOR INVESTIGACIÓ EN SALUT PÚBLICA · VNIVERSITAT DE VALÈNCIA · ciberesp

Validation and accreditation

17 The following wet lab parameters in the validation process should be included in the validation: sample type, sample volume, extraction protocol, library preperation protocol.

18 The following sequencing parameters should be included in the validation process: precision, accuracy of sequence output, sequence depth, analytical sensitivity, specificity, limit of detection.

19 Result interpretation: a cut-off for defining a positive result (read count, coverage) should be determined based on validation data, e.g. comparison with PCR results, using prototype viruses. For defining a positive result, use a threshold of three distinctly covered genome regions after background subtraction based on negative controls.

20 An external quality assessment programs (EQA) should be adhered to evaluate the performance of metagenomics protocols applied in diagnostic settings, assessing both qualitative (correct pathogen detection) and quantitative characteristics (target read numbers).

Validation for each sample type: volume, extraction, library prep.

Sequencing: precision, sensitivity, specificity, LOD

		Clinical Dx* (direct detection CSF testing) Pos	Clinical Dx* (direct detection CSF testing) Neg
mNGS	Pos	32	3
	Neg	8	164
		sensitivity = 80.0%	*PPV = 91.4%*
		specificity = 98.2%	***NPV = 95.3%***

		Clinical Dx* (direct detection CSF testing) Pos	Clinical Dx* (direct detection CSF testing) Neg
Conventional testing (direct detection CSF testing)	Pos	27	1#
	Neg	13	164
		sensitivity = 67.5%	*PPV = 96.4%*
		specificity = 99.4%	***NPV = 92.7%***

Wilson *et al.*, 2019 *N Eng J Med* **380**:2327-2340

ATCC · QCMD

Metagenomic Control Material (bacteria)

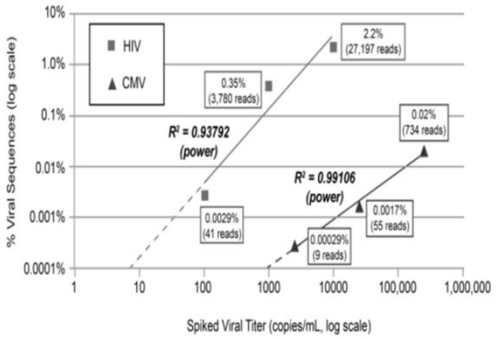
Viral Metagenomics NGS EQA (Q4 - Pilot Study)

ESCV European Society for Clinical Virology — ENNGS ESCV NETWORK ON NEXT-GENERATION SEQUENCING

GENERALITAT VALENCIANA · Fundació Fisabio · CENTRE SUPERIOR INVESTIGACIÓ EN SALUT PÚBLICA · VNIVERSITAT DE VALÈNCIA · ciberesp

F. Xavier López-Labrador – XXXI Jornades SCMIMC, Montbrió del Camp 15/7/2022

Validation: sensitivity, cutoff, threshold, contamination



Miller et al. Genome Research, 2019, 29(5):831-84.

RNA viruses detected to <100 copies/ml
(~ Real-time PCR)

DNA viruses detected to ~1000 copies/ml
(100-fold < Real-time PCR)
(~ 16S pan-bacterial PCR)

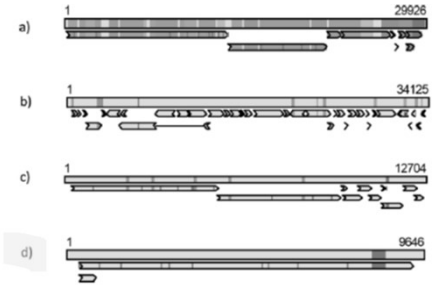


Fig. 1. Examples of coverage plots [46] with rue positive mNGS findings (a–c) confirmed by PCR in real clinical samples: a) human coronavirus HKU-1, 3951 reads, 89 % genome coverage, b) human mastadenovirus A, 19 reads, 8% genome coverage, >3 genome locations, and c) spiked-in equine arteritis virus, 14 reads, 5% genome coverage >3 genome locations, and d) an example of a false positive mNGS finding plotting a mapped hepatitis C virus amplicon contaminant, 133,213 reads, 4% coverage but only 1 genome location.
Top bar represents nucleotide alignment, bottom bar(s) represents amino acid alignment, green zone: matching sequences.
Distribution of reads over the genome is an important parameter for defining a positive result.

de Vries et al. J Clin Virol 2021

Journal of Clinical Virology 138 (2021) 104812

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

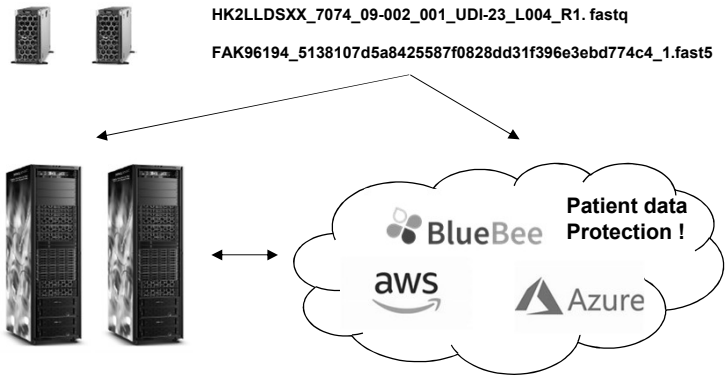
Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting

Jutte J.C. de Vries [a,*], Julianne R. Brown [b], Natacha Couto [c], Martin Beer [d], Philippe Le Mercier [e], Igor Sidorov [a], Anna Papa [f], Nicole Fischer [g], Bas B. Oude Munnink [h], Christophe Rodriquez [i], Maryam Zaheri [j], Arzu Sayiner [k], Mario Hönemann [l], Alba Perez Cataluna [m], Ellen C. Carbo [a], Claudia Bachofen [n], Jakub Kubacki [n], Dennis Schmitz [o], Katerina Tsioka [f], Sébastien Matamoros [p], Dirk Höper [d], Marta Hernandez [q], Elisabeth Puchhammer-Stöckl [r], Aitana Lebrand [e], Michael Huber [j], Peter Simmonds [s], Eric C.J. Claas [a], F. Xavier López-Labrador [t,u,v,**], on behalf of the ESCV Network on Next-Generation Sequencing

F. Xavier López-Labrador – XXXI Jornades SCMIMC, Montbrió del Camp 15/7/2022

IT equipment, software, data security and storage

1. Given the amount of data and pipelines for metagenomic analysis, the use of a cluster server, usually situated within a dedicated physically separated "core" IT infrastructure facility for central data processing facilities is recommended, either accessible directly or via external providers of the analysis pipelines.
2. It is recommended to have written agreements with cloud service providers on the management of protection information for unauthorized access, use, disclosure, disruption, modification, or destruction, confidentiality and timely/reliable access to and use of information. The agreement should also include the management of new releases of software versions to enable validation prior to using a new version for patient care.
3. NGS FASTQ data and metadata files should be stored with file names and folders having unique and identifying names helpful in classifying and sorting (https://www.ukdataservice.ac.uk/manage-data/format/organising.aspx)



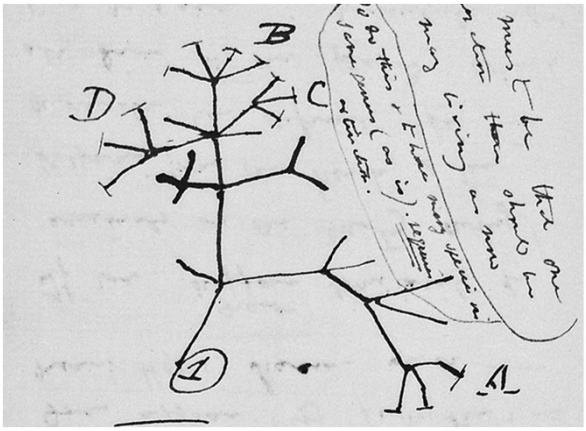
ESCV European Society for Clinical Virology — ENNGS ESCV NETWORK ON NEXT-GENERATION SEQUENCING

GENERALITAT VALENCIANA — Fundació Fisabio — CENTRE SUPERIOR INVESTIGACIÓ EN SALUT PÚBLICA — VNIVERSITAT DE VALÈNCIA — ciberesp

Version control (pipeline & database)

5. The reference database should consist of genomes that cover the entire genetic diversity of relevant organisms and should be curated in order not to contain any artificial, low-quality or incorrectly named genome sequences.

6. It is recommended to periodically update the reference databases used for taxonomic profiling, and to validate this update. The frequency of the update is dependent on the need to classify at subtype or isolate level, and on the appearance of novel viruses in the updated public databases.



ESCV European Society for Clinical Virology — ENNGS ESCV NETWORK ON NEXT-GENERATION SEQUENCING

GENERALITAT VALENCIANA — Fundació Fisabio — CENTRE SUPERIOR INVESTIGACIÓ EN SALUT PÚBLICA — VNIVERSITAT DE VALÈNCIA — ciberesp

Journal of Clinical Virology 141 (2021) 104908

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples

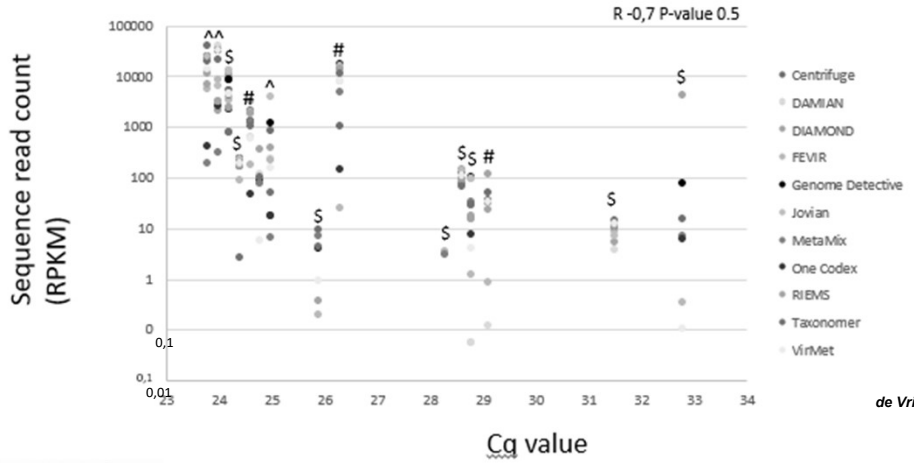
Jutte J.C. de Vries [a,*], Julianne R. Brown [b], Nicole Fischer [d], Igor A. Sidorov [a], Sofia Morfopoulou [c], Jiabin Huang [d], Bas B. Oude Munnink [e], Arzu Sayiner [f], Alihan Bulgurcu [g], Christophe Rodriguez [g], Guillaume Gricourt [g], Els Keyaerts [h], Leen Beller [h], Claudia Bachofen [i], Jakub Kubacki [i], Cordey Samuel [j], Laubscher Florian [j], Schmitz Dennis [k], Martin Beer [l], Dirk Hoeper [l], Michael Huber [m], Verena Kufner [m], Maryam Zaheri [m], Aitana Lebrand [n], Anna Papa [o], Sander van Boheemen [e], Aloys C.M. Kroes [a], Judith Breuer [b,c], F. Xavier Lopez-Labrador [p,q], Eric C.J. Claas [a]

Benchmarking: datasets

- 13 clinical metagenomic datasets, well-characterized samples by PCR
- Patients with encephalitis or respiratory complaints
  - CSF (n=4)
  - Brain biopsies (n=3)
  - Nasopharyngeal swabs (n=3)
  - Nasal washings (n=1)
  - Bronchoalveolar lavage (n=1)
  - Plasma (n=1)

F. Xavier López-Labrador – XXXI Jornades SCMIMC, Montbrió del Camp 15/7/2022

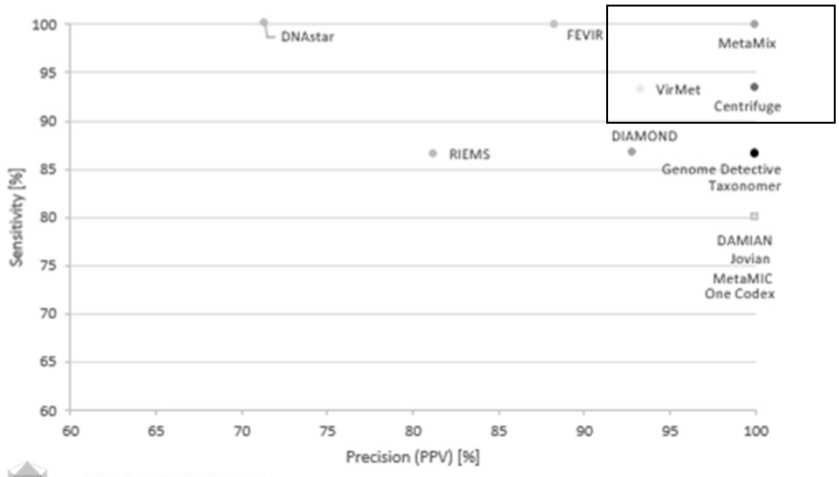
Semi-quanitative results, sensitivity



de Vries et al. J Clin Virol 2021

[^ mRNA sequencing, $ RNA/DNA sequencing, and #: a captured approach using probes targeting vertebrate viruses

Overall score (sensitivity/PPV)



	PCR+	PCR -	
mNGS +	TP	FP	PPV
mNGS -	FN	TN?	NPV
	Sensitivity	Specificity	

de Vries et al. J Clin Virol 2021

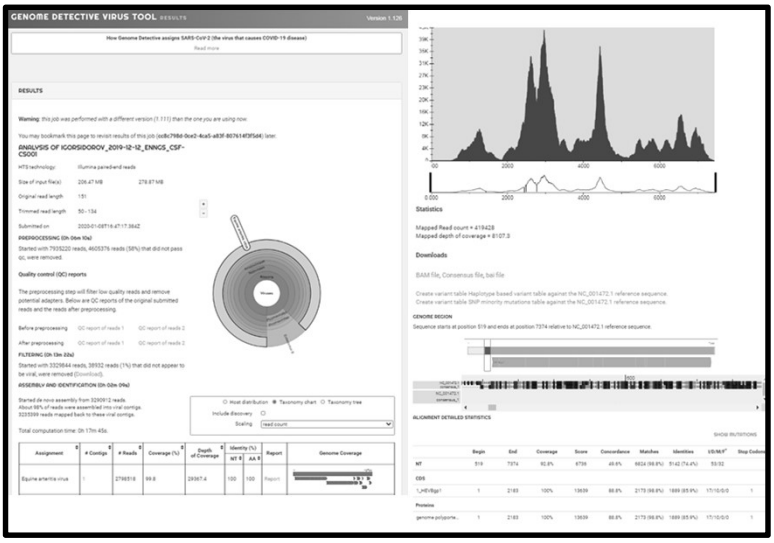
Result review and reporting

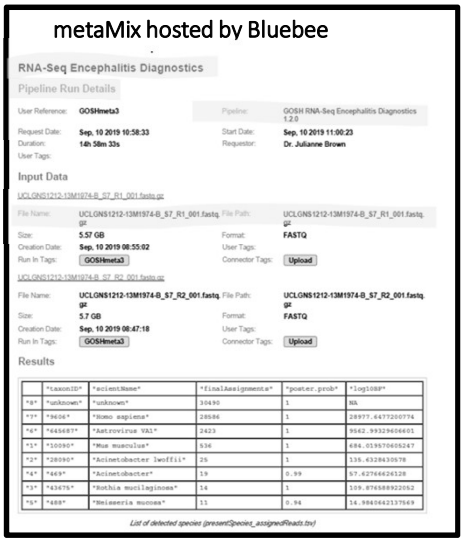
12. Before reporting, the mNGS data need to be technically evaluated and reviewed, for quality, possible laboratory contaminations and plausibility.
13. Hits of known reagent contaminants, misassignments, bacteriophages, and common (retro)viral endogenous sequences should not be reported to the clinician.

The result of mNGS should be reported to the clinician in a compact format and facilitate decision making

The reports should be comprehensible, but yet easy to read and contain only clinically relevant or potentially relevant information.

User-friendly output formats





Conclusions

- For some clinical syndromes, such as encephalitis, there is a need to extend the diagnostic portfolio with mNGS.
- For many others, cost and turn-around-time constraints preclude mNGS of completely replacing conventional diagnostic testing in the near future.
- Technical, procedural and financial parameters will develop rapidly: future developments will support the progressive and broad introduction of clinical metagenomic sequencing.
- Bioinformatic software tools and platforms will develop very fast, which will support the progressive and broad introduction of metagenomic sequencing into Clinical Microbiology and Public Health laboratories
- ESCV recommendations are intended to guide laboratories on the implementation of mNGS and bioinformatics diagnostic workflows.
- Future: Real-time sequencing and simultaneous transcriptome analysis in development

ENNGS network and collaborators

ESCV European Society for Clinical Virology

ENNGS ESCV NETWORK ON NEXT-GENERATION SEQUENCING

Jutte de Vries (NL, Leiden, LUMC)
Claudia Bachofen and Jakub Kubacki (SW, Zurich, UZ)
Samuel Cordey and Florian Laubscher (SW, Geneva, UHG)
Dennis Schmitz (NL, Bilthoven, RIVM)
Martin Beer and Dirk Hoeper (GE, Greifswald, FLI)
Michael Huber, Verena Kufner and Maryam Zaheri (SW, Zurich, UZ)
Aitana Lebrand (SW, Geneva, SIB)
Anna Papa (GR, Thessaloniki, AUT)
F. Xavier López-Labrador (ES, FISABIO-Public Health)
Javier Buesa and Jesús Rodríguez (ES, Univ. Valencia)

Charles Y. Chiu and Steve Miller (USA, UCSF VDDC)

Julianne R. Brown, Sofia Morfopoulou and Judith Breuer (UK, London, GOSH)

Nicole Fischer and Jiabin Huang (GE, Hamburg, UMCH-E)

Igor A. Sidorov, Eric C.J. Claas and Aloys Kroes (NL, Leiden, LUMC)

Bas B. Oude Munnink and Sander van Boheemen (NL, Rotterdam, EMC)

Arzu Sayiner and Alihan Bulgurcu (TU, Izmir, DEU)

Christophe Rodriguez and Guillaume Gricourt (FR, Paris, HHM)

Els Keyaerts and Leen Beller (BE, Leuven, KUL)

F. Xavier López-Labrador – XXXI Jornades SCMIMC, Montbrió del Camp 15/7/2022

ACKNOWLEDGMENTS

Virology Lab, Genomics and Health Area FISABIO-Public Health

Beatriz Mengual Chulià
Laura Cano
Alessandra Franze'
Sandra García
FX López-Labrador

Funding / support:

AES PI18/0159 - PI18/01821

Valencia Hospital Network for the Study of Influenza and Other Respiratory Viruses (VAHNSI)

Site's PIs & collaborators:

M. Carballido-Fernández
J. Mollar-Maseres
M. Tortajada-Girbés
G. Schwarz-Chávarri

Field researchers:

Ángela Pla
Cristina Mestre
Beatriz Segura
Victoria Ruiz

Vaccine Research Area, FISABIO-Public Health

Coordination Office:

A. Orrico
A. Mira-Iglesias
C. Vergara-Hernández
A. García-Andreu
J. Díez-Domingo