Review Article

Metagenomics-driven Virome: Current Procedures and New Additions

Claudia Kohl*, Andreas Nitsche, Andreas Kurth

Robert Koch Institute, Centre for Biological Threats and Special Pathogens, Seestrasse 10, 13353 Berlin, Germany

Abstract | Next generation sequencing (NGS) opened up a plethora of new research possibilities in biology and medicine. Metagenomics is one of these emerging NGS applications and offers the opportunity to study i.e. whole ecosystems. Basically, the metagenomics approach is similar to well-known shotgun-sequencing, though on a much bigger scale. For instance, the metagenome of a lake would include all the fish, ducks, plants, fungi, bacteria and everything else that belongs to the lake. If we apply this approach to clinical samples we can identify the community of etiological pathogens, without any knowledge on the targets in advance. However, clinical specimens usually comprise an overwhelming amount of host nucleic acids, which by far exceeds the number of pathogen nucleic acids in the sample. Subsequently, it is necessary to either decrease the amount of host nucleic acids or increase the amount of pathogen nucleic acids, to allow for detection via metagenomic NGS. This minireview is revising our developed TUViD-VM protocol and selected other approaches regarding their suitability in metagenomics. We provide an overview on the difficulties, challenges and opportunities that developed alongside metagenomic virus discovery. The field of metagenomics from clinical specimens promises the identification of novel, yet unknown, infectious diseases and etiologies.

Editor | Muhammad Munir, The Pirbright Institute, UK.
Received | October 6, 2015; **Accepted** | December 12, 2015; **Published** | January 13, 2016
***Correspondence** | Claudia Kohl, Centre for Biological Threats and Special Pathogens, Germany; **Email:** kohlc@rki.de
DOI | http://dx.doi.org/10.17582/journal.bjv/2015.2.6.96.101
Citation | Kohl C., A. Nitsche and A. Kurth. 2015. Metagenomics-driven virome: current procedures and new additions. *British Journal of Virology*, 2(6): 96-101.

Metagenome vs. Virus Discovery

When talking about metagenomics we usually think about the analysis of microbial communities in a selected environment, like bacteria in soil or in the human gut. The term 'metagenome' is built from meta-analysis (the statistical approach to normalize quantifiably data from differing sources) and genome (the total genetic material of an organism). The metagenome-based approach promises to represent quantifiably the ratios of different phyla within a selected group, which could be compared to the ratios of phyla within another group (e.g. microbial communities in the sediments of two different lakes). Subsequently, it aims at comparing quantifiably, by definition, spatial or ecologically differing habitats: for instance the comparison of different saline water habitats regarding microbial diversity (Siddhapura et al., 2010). The beauty of this methodology, especially if combined with next generation sequencing (NGS), was quickly also recognized by microbe hunters around the world (Bibby, 2013; Carrington, 2012; Edwards and Rohwer, 2005; Forde and O'Toole, 2013; Fricke et al., 2009; Radford et al., 2012; Simon and Daniel, 2011; Svraka et al., 2010; Tang and Chiu, 2010; Thurber et al., 2009; Wooley et al., 2010). Unlike most other hunting techniques, a metagenome does not require any knowledge about the prey in advance. For bacterial metagenomics, 16S rRNA amplicon sequencing opened up a new field in functional and

ecological microbe detection and analysis (Fierer et al., 2010; Qin et al., 2010).

In contrast, when aiming at virus discovery, the metagenome approach is not equally auspicious. Comparing bacteria and viruses, we find significant differences that can constitute drawbacks for virus detection:

- Viruses are much smaller than bacteria and replicate within the host cells; separation of viruses from their host cells is therefore much harder.
- Viruses do not share a common feature, like 16S rRNA or a similar region that could be amplified with 'Pan-virus primers'.
- Virus genomes are much smaller than bacterial genomes – smaller genomes = less fragments in NGS preparation and detection is less likely.

Subsequently, applying metagenomics to virus detection requires taking a small step back from the initial idea of metagenomics. The *virome* (an approach only looking for viruses using metagenomic protocols) is therefore a smaller and more limited version of a *real* metagenome.

About ratios

The chances of detecting viruses by metagenomic NGS approaches rise and fall with the ratio between the nucleic acid of a virus and the background in a given sample. To stay with the hunting example: if we are looking for five rabbits in a deep forest full of other game it is nearly impossible to track them, whereas five rabbits in the open field are easy to find. Metagenomic NGS does not distinguish between interesting and uninteresting nucleic acids, thus the sequencing result will always depend on the ratio of interesting to uninteresting sequences in the sample. The advantage of this technique is at the same time its disadvantage: Every sequence present in the sample will be sequenced simultaneously. Here one needs to take into consideration that the amount of host-nucleic acids far exceeds the amount of virus-nucleic acids per cell. A single human cell contains numerous amounts of nucleic acids: 3.27×10^9 bp of genomic DNA and a plethora of different RNA species (Venter et al., 2001). If the same cell is infected with a virus the viral nucleic acids usually contribute less than one per mill to the total amount of nucleic acid. Moreover, viruses replicate to varying yields in different tissues, which may lead to even less viral nucleic acids when investigating a non-optimal tissue. Finally, there are two obvious solutions to get more sequence information of interest:

1. Increase the amount of interesting sequence information (e.g. virus propagation in cell culture).
2. Decrease the amount of uninteresting sequence information

Nowadays, various NGS approaches already provide reliable solutions for the first option (Kohl et al., 2012a, 2012b; Radonić et al., 2014; Svraka et al., 2010), but when it comes to clinical specimens like blood, fluids or even infected organ tissue, the successful detection of viruses is possible, but much less likely, and it is necessary to think about the second option. Using tissue for virus detection allows for the elucidation of viral infections directly at the site of viral replication. This, in turn, allows for the instant correlation of physiological host effects (*phenotype*) with the causing viral agent (*genotype*). Clinical specimens other than tissue are mainly host excretions (e.g. urine or blood), and their viral load is therefore dependent on transportation fluids and often less concentrated. The availability of viruses in infected organ tissue is less dependent on stages and cycles of replication, viremia and shedding, respectively. Even though the detection of viruses directly from infected organ tissue offers obvious and valuable advantages, only few studies have used this approach. On the other hand, organ tissue is usually not easily available or requires invasive techniques, beside the crucial question of picking the right organ. Reliable virus purification from tissue remains a challenge.

Indeed, bioinformatic pathogen analysis pipelines are available and are promising rapid and reliable identification. However, the ratio of viruses to host-genome is still critical. A bad ratio requires a great sequencing depth to be able to identify enough sequences of interest: this is time and cost consuming. The comparison of different analysis pipelines is complicated and depends on the respective research question (Baker et al., 2013, Naccache et al., 2014).

Protocols

Researchers have published protocols for sequencing

of virus particles from different sources (i.e. soil, blood, tissue, plants and liquids) (Alavandi and Poornima, 2012; Culley et al., 2006; Djikeng et al., 2009; Radonić et al., 2014; Sachsenröder et al., 2012; Tang and Chiu, 2010; Thurber et al., 2009; Whon et al., 2012; Winget and Wommack, 2008). Whichever protocol is used, it is of major importance that the sample is as native as possible or deeply frozen until preparation and that the viral capsids or nucleocapsids are still intact to allow for a successful separation. The general purification procedure for clinical tissue specimens is summarized here:

The first step of purification is the disruption of the tissue/cells or aggregates for the release of viral particles; some protocols use bead mills or vortexes and others use shredder kits (Donaldson et al., 2010; Ge et al., 2012; Li et al., 2010; Miranda and Miranda, 2011; Phillips et al., 2012; Victoria et al., 2009). However, host-nucleic acids, proteins and cell organelles will be released simultaneously, and a strategy is necessary to enrich viral particles preferably, while decreasing the host genome. The mechanical methods of choice are ultra-centrifugation, filtration and tangential flow (Kohl et al., 2012b; Potgieter et al., 2009; Sachsenröder et al., 2012; Victoria et al., 2009). These three methods again have numerous variations, like using sucrose or caesium chloride for ultra-centrifugation, filter size and speed of tangential flow. Often the host-genome is decreased additionally by enzymatic digestion (Donaldson et al., 2010; Li et al., 2010; Victoria et al., 2009). The viral nucleic acids are protected in the capsid, and only the surrounding uninteresting host sequences are digested. In the majority DNase treatment is performed, sometimes also RNase and Benzonase are utilized.

At this stage, the sample is ready for RNA and DNA extraction. Various protocols describe several successful methods for extraction of nucleic acids, ranging from classical phenol-chloroform to high-throughput kits. Following extraction, the viral nucleic acids may need to be generically amplified to increase the detectability. Published protocols often use the K-primers, DOP, commercial kits or other random techniques to amplify viral nucleic acids (Cheval et al., 2011; Nanda et al., 2008; Stang and Korn, 2005; Telenius et al., 1992; Uhlenhaut et al., 2009).

The published protocols are successful in detecting particular viruses. When we looked for a general protocol for virus purification we found it hard to compare all the different approaches. This has been the reason for the development of the TUViD-VM protocol (Kohl et al., 2015).

TUViD-VM

To develop the TUViD-VM protocol, every single purification step was compared to a set of commonly used purification methods and was further evaluated in order to result in maximum likelihood virus detection for four different model viruses (Kohl et al., 2015). First we designed a comparable tissue model based on internal organs of chicken, each infected with one out of four viruses at low concentrations. These viruses were chosen based on their significance in the context of emerging zoonotic diseases and on their morphological and molecular heterogeneity to obtain results for a broad range of viruses. This final protocol was validated and adjusted until minimal host nucleic acids were detected by qPCR while maximising the amount of the viral nucleic acids amplified. We finally validated the protocol by next generation sequencing and confirmed the qPCR results. We applied TUViD-VM to a clinical sample and confirmed our findings again. In the end, we reduced the forest and made the rabbits detectable.

The TUViD-VM publication is of interest for researchers looking for a straightforward protocol for viral metagenomics from tissue samples and for those interested in performance of the single purification methods used and their effects on the detectability of different viruses. The protocol was developed with a panel of four defined viruses for which it worked well. It is possible that the protocol is therefore slightly biased toward the detection of these virus types (reovirus, paramyxovirus, poxvirus, influenza virus), although we have chosen viruses that were as different as possible regarding e.g. capsid structure, genome orientation, sensitivity and density. For each set of purification methods (e.g., homogenization) the results are displayed for all tested approaches and viruses (Kohl et al., 2015). The presented methodology is also of benefit for researchers just looking for a simple technique not restricted to particular virus detection – as for virome studies or outbreak investigations where knowledge of the viruses is not available in advance. However, the purification of viruses from tissue using TUViD-VM represents only the first, albeit important, step of virus identification. The second step, the

bioinformatic identification, is at least as crucial as the purification itself.

For a perspective ahead we may go back to the metagenome approach from the very beginning. If we want to draw spatial and ecological conclusions also from our virome samples we need to reestablish comparable conditions. The difficulty here is that viruses need to be purified for their sole detection whereas bacteria can be detected straight away. Purification shifts the ratios of detectable nucleic acids. To use the same for ecological virome approaches would allow adding more meaning to virus-hunting and virome studies could be lifted to the next level.

Acknowledgments

The authors are grateful to Ursula Erikli for copy-editing.

Authors Contributions

All authors reviewed the literature and wrote and discussed the manuscript.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

- Alavandi, S. V., Poornima, M. Viral Metagenomics: A Tool for Virus Discovery and Diversity in Aquaculture. Indian J Virol, 2012; 23(2):88–98. http://dx.doi.org/10.1007/s13337-012-0075-2
- Bibby, K. Metagenomic identification of viral pathogens. Trends Biotechnol, 2013; 31(5): 275–9. http://dx.doi.org/10.1016/j.tibtech.2013.01.016
- Baker, K. S., Leggett, R. M., Bexfield, N. H., Alston, M., Daly, G., Todd, S., Tachedjian, M., Holmes, C. E., Crameri, S., Wang, L. F., Heeney, J. L., Suu-Ire, R., Kellam, P., Cunningham, A. A., Wood, J. L., Caccamo, M., Murcia, P. R. Metagenomic study of the viruses of African straw-coloured fruit bats: detection of a chiropteran poxvirus and isolation of a novel adenovirus. Virology, 2013; 441(2): 95–106. http://dx.doi.org/10.1016/j.virol.2013.03.014
- Carrington, C. V. F., 2012: Viral Genomics: Implications for the Understanding and Control of Emerging Viral Diseases. In: Nelson, K E, Jones-Nelson, B (eds), Genomics Applications for the Developing World, pp. 91–114. Springer.
- Cheval, J., Sauvage, V., Frangeul, L., Dacheux, L., Guigon, G., Dumey, N., Pariente, K., Rousseaux, C., Dorange, F., Berthet, N., et al. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. J Clin Microbiol, 2011; 49(9): 3268–75. doi: 10.1128/JCM.00850-11.
- Culley, A. I., Lang, A. S., Suttle, C. A. Metagenomic analysis of coastal RNA virus communities. Science, 2006; 312(5781): 1795–8. doi: 10.1126/science.1127404.
- Djikeng, A., Kuzmickas, R., Anderson, N. G., Spiro, D. J. Metagenomic analysis of RNA viruses in a fresh water lake. PLoS One, 2009; 4(9): e7264. doi: 10.1371/journal.pone.0007264.
- Donaldson, E. E. F., Haskew, A. N. A., Gates, J. E., Huynh, J., Moore, C. J., Frieman, M. B. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. J Virol, 2010; 84(24): 13004–18. doi: 10.1128/JVI.01255-10.
- Edwards, R. A., Rohwer, F. Viral metagenomics. Nat Rev Microbiol, 2005; 3(6): 801–5. doi:10.1038/nrmicro1163.
- Fierer, N., Lauber, C.L., Zhou, N., McDonald, D., Costello, E. K., Knight, R. Forensic identification using skin bacterial communities. Proc Natl Acad Sci U S A, 2010; 107(14): 6477–81. doi: 10.1073/pnas.1000162107.
- Forde, B. M., O'Toole, P. W. Next-generation sequencing technologies and their impact on microbial genomics. Brief Funct Genomics, 2013; 12(5): 440–53. doi: 10.1093/bfgp/els062.
- Fricke, W. F., Rasko, D. A., Ravel, J. The role of genomics in the identification, prediction, and prevention of biological threats. PLoS Biol, 2009; 7(10): e1000217. doi:10.1371/journal.pbio.1000217.
- Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y., Shi, Z. Metagenomic Analysis of Viruses from the Bat Fecal Samples Reveals Many Novel Viruses in Insectivorous Bats in China. J Virol, 2012; 86(8): 4620–30. doi: 10.1128/JVI.06671-11.
- Kohl, C., Vidovszky, M. Z., Mühldorfer, K., Dabrowski, P. W., Radonić, A., Nitsche, A., Wibbelt, G., Kurth, A., Harrach, B. Genome analysis of bat

adenovirus 2: indications of interspecies transmission. J Virol, 2012a; 86(3): 1888–92. doi: 10.1128/JVI.05974-11.

- Kohl, C., Lesnik, R., Brinkmann, A., Ebinger, A., Radonić, A., Nitsche, A., Mühldorfer, K., Wibbelt, G., Kurth, A. Isolation and characterization of three mammalian orthoreoviruses from European bats. PLoS One, 2012b; 7(8): e43106. doi: 10.1371/journal.pone.0043106.
- Kohl, C., Brinkmann, A., Dabrowski, P.W., Radonić, A., Nitsche, A., Kurth, A. Protocol for metagenomic virus detection in clinical specimens. Emerg Infect Dis, 2015; 21(1): 48–57. doi: 10.3201/eid2101.140766.
- Li, L., Victoria, J. G., Wang, C., Jones, M., Fellers, G. M., Kunz, T.H., Delwart, E. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J Virol, 2010; 84(14): 6955–65. doi: 10.1128/JVI.00501-10.
- Miranda, M. E. G., Miranda, N. L. J. Reston ebolavirus in humans and animals in the Philippines: a review. J Infect Dis, 2011; 204 Suppl 3, S757–60. doi: 10.1093/infdis/jir296.
- Nanda, S., Jayan, G., Voulgaropoulou, F., Sierra-Honigmann, A. M., Uhlenhaut, C., McWatters, B. J. P., Patel, A., Krause, P. R. Universal virus detection by degenerate-oligonucleotide primed polymerase chain reaction of purified viral nucleic acids. J Virol Methods, 2008; 152(1–2): 18–24. doi: 10.1016/j.jviromet.2008.06.007.
- Naccache, S. N., Federman, S., Veeraraghavan, N., Zaharia, M., Lee, D., Samayoa, E., Bouquet, J., Greninger, A. L., Luk, K. C., Enge, B., Wadford, D. A., Messenger, S. L., Genrich, G. L., Pellegrino, K., Grard, G., Leroy, E., Schneider, B. S., Fair, J. N., Martínez, M. A., Isa, P., Crump, J. A., DeRisi, J. L., Sittler, T., Hackett, J. Jr., Miller, S., Chiu, C. Y.. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res, 2014; 24(7): 1180–92. doi: 10.1101/gr.171934.113.
- Phillips, C. D., Phelan, G., Dowd, S. E., McDonough, M. M., Ferguson, A. W., Delton Hanson, J., Siles, L., Ordóñez-Garza, N., San Francisco, M., Baker, R. J. Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. Mol Ecol, 2012; 21(11): 2617–27. doi: 10.1111/j.1365-294X.2012.05568.x.
- Potgieter, A. C., Page, N. A., Liebenberg, J., Wright, I. M., Landt, O., van Dijk, A. A. Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. J Gen Virol, 2009; 90(Pt 6): 1423–32. doi: 10.1099/vir.0.009381-0.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature, 2010; 464(7285): 59–65. doi: 10.1038/nature08821.
- Radford, A. D., Chapman, D., Dixon, L., Chantrey, J., Darby, A. C., Hall, N. Application of next-generation sequencing technologies in virology. J Gen Virol, 2012; 93(Pt 9): 1853–68. doi: 10.1099/vir.0.043182-0.
- Radonić, A., Metzger, S., Dabrowski, P.W., Couacy-Hymann, E., Schuenadel, L., Kurth, A., Mätz-Rensing, K., Boesch, C., Leendertz, F. H., Nitsche, A. Fatal monkeypox in wild-living sooty mangabey, Côte d'Ivoire, 2012. Emerg Infect Dis, 2014; 20(6): 1009–11. doi: 10.3201/eid2006.13-1329.
- Sachsenröder, J., Twardziok, S., Hammerl, J. A, Janczyk, P., Wrede, P., Hertwig, S., Johne, R. Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. PLoS One, 2012; 7(4): e34631. doi: 10.1371/journal.pone.0034631.
- Siddhapura, P. K., Vanparia, S., Purohit, M. K., Singh, S. P. Comparative studies on the extraction of metagenomic DNA from the saline habitats of Coastal Gujarat and Sambhar Lake, Rajasthan (India) in prospect of molecular diversity and search for novel biocatalysts. Int J Biol Macromol, 2010; 47(3): 375–9. doi: 10.1016/j.ijbiomac.2010.06.004.
- Simon, C., Daniel, R. Metagenomic analyses: past and future trends. Appl Environ Microbiol, 2011; 77(4): 1153–61. doi: 10.1128/AEM.02345-10.
- Stang, A., Korn, K., Wildner, O., Uberla, K. Characterization of virus isolates by particle-associated nucleic acid PCR. J Clin Microbiol, 2005; 43(2): 716–20. doi: 10.1128/JCM.43.2.716-720.2005.
- Svraka, S., Rosario, K., Duizer, E., van der Avoort, H., Breitbart, M., Koopmans, M. Metagenomic sequencing for virus identification in a public-health setting. J Gen Virol, 2010; 91(Pt 11): 2846–56. doi: 10.1099/vir.0.024612-0.
- Tang, P., Chiu, C. Metagenomics for the discovery

of novel human viruses. Future Microbiol, 2010; 5(2): 177–89. doi: 10.2217/fmb.09.120.

- Telenius, H., Carter, N. P., Bebb, C. E., Nordenskjöld, M., Ponder, B. A. J., Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. Genomics 13(3): 718–25. doi:10.1016/0888-7543(92)90147-K.
- Thurber, R. V, Haynes, M., Breitbart, M., Wegley, L., Rohwer, F. Laboratory procedures to generate viral metagenomes. Nat Protoc, 2009; 4(4): 470–83. doi: 10.1038/nprot.2009.10.
- Uhlenhaut, C., Cohen, J. I., Fedorko, D., Nanda, S., Krause, P. R. Use of a universal virus detection assay to identify human metapneumovirus in a hematopoietic stem cell transplant recipient with pneumonia of unknown origin. J Clin Virol, 2009; 44(4): 337–9. doi: 10.1016/j.jcv.2009.01.011.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. The sequence of the human genome. Science, 2001; 291(5507): 1304–51. doi: 10.1126/science.1058040.
- Victoria, J.G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S., Delwart, E. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol, 2009; 83(9): 4642–51. doi: 10.1128/JVI.02301-08.
- Whon, T.W., Kim, M.-S., Roh, S.W., Shin, N.-R., Lee, H.-W., Bae, J.-W. Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere. J Virol, 2012; 86: 8221–31. doi: 10.1128/JVI.00293-12.
- Winget, D.M., Wommack, K.E. Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. Appl Environ Microbiol, 2008: 74(9): 2612–8. doi: 10.1128/AEM.02829-07.
- Wooley, J.C., Godzik, A., Friedberg, I. A primer on metagenomics. PLoS Comput Biol, 2010; 6(2): e1000667. doi: 10.1371/journal.pcbi.1000667.