

Silvana Beres Castrignano
Teresa Keico Nagasse-Sugahara

The metagenomic approach and causality in virology

Abordagem metagenômica e causalidade em virologia

ABSTRACT

Nowadays, the metagenomic approach has been a very important tool in the discovery of new viruses in environmental and biological samples. Here we discuss how these discoveries may help to elucidate the etiology of diseases and the criteria necessary to establish a causal association between a virus and a disease.

DESCRIPTORS: Metagenomics. Viruses. Virology. Communicable Diseases, etiology. Causality. Koch's Postulates.

RESUMO

A abordagem metagenômica tem sido ferramenta muito importante atualmente na descoberta de novos vírus em amostras ambientais e biológicas. Aqui discutimos a maneira como essas descobertas podem ajudar a elucidar a etiologia de doenças e os critérios necessários para que a associação causal entre um vírus e uma doença seja estabelecida.

DESCRIPTORIOS: Metagenômica. Vírus. Virologia. Doenças Transmissíveis, etiologia. Causalidade. Postulados de Koch.

Núcleo de Doenças Respiratórias. Centro de Virologia. Instituto Adolfo Lutz. São Paulo, SP, Brasil

Correspondence:

Silvana Beres Castrignano
Centro de Virologia - Instituto Adolfo Lutz
Av. Dr. Arnaldo, 355
01246-902 São Paulo, SP, Brasil
E-mail: s.castrignano@hotmail.com

Received: 3/28/2014
Approved: 8/7/2014

Article available from: www.scielo.br/rsp

INTRODUCTION

Before the association between an infectious agent and a disease is made public, it needs to be well established. The adequate treatment of patients suffering from an infectious disease, implementation of preventive measures, understanding of the different phases of the disease, development of the therapies, and occasionally, the development of a vaccine depend on the elucidation of this association.¹³ The criteria for achieving correct associations are being proposed by eminent scientists.^{7,9-11,13,16,19}

With regard to viruses, which are the focus here, the need for establishing a causal association is a very contemporary subject because many viruses are now being discovered. Metagenomic techniques and next-generation sequencing platforms are the main drivers of these discoveries. At the same time, several potentially infectious diseases and several cases of common infectious diseases, including acute respiratory disease, encephalitis, acute gastroenteritis, and hepatitis, do not have a known etiological agent.^{12,16}

CRITERIA FOR CAUSAL ASSOCIATION BETWEEN AGENTS AND DISEASES

In the end of the 19th century, Koch¹¹ defended the fundamental principles that helped establish the causal association between microorganisms and infectious diseases according to a set of criteria laid out by him. According to this author, a proof of causality requires the agent to be present in every case of a particular disease, to be absent in other diseases, and after isolation and culture, be sufficient to reproduce the disease by inoculation of a susceptible host.

Since then, some modifications of these postulates have been proposed, including the following: the acknowledgment of viruses as infectious agents (they were unknown in Koch's time),^{7,19} of the importance of the study of antibodies (presence and time of appearance),¹⁹ of the possibility of disease prevention using vaccines against the virus,¹⁰ of the importance of epidemiological studies,¹⁰ and of the concept that several factors, and not only a single cause, can contribute to disease development.⁷

Owing to the advances in molecular biology techniques, proposals to amend Koch's postulates from the 80s onwards included criteria based on microbial genetics.^{9,13,16} Among these recent propositions, both the criteria of Mokili et al,¹⁶ based on the comparison of metagenomic characteristics among infected and healthy individuals, and the criteria of Lipkin,¹³ who grouped laboratory, clinical and epidemiological data into three certainty levels to establish an association between pathogens and diseases, considered the inoculation of the infectious agent in a healthy individual as

a criterion for the confirmation of causality (this criterion was inherited from Koch's postulates). However, because of ethical issues, it is not possible to inoculate a suspected pathogen in human beings, and with exceptions such as the SARS-related human coronavirus, few etiological agents have a susceptible experimental animal model.² Moreover, Lipkin¹³ suggested an alternative to this rule from Koch's postulates and indicated that a causal association can be confirmed if the disease can be attenuated or prevented with the use of microorganism-specific vaccines, drugs, or antibodies.

On the other hand, according to the molecular guidelines of Fredericks & Relman⁹ for establishing microbial disease causation, there is no need to isolate a virus or inoculate it in a host. These guidelines are based on the identification of the microbial genome using *in situ* hybridization in tissue samples with pathological changes and on the analysis of the copy number of pathogen-associated nucleic acid sequences in tissue samples with or without lesions during several phases of the disease, including the period before disease onset.

Because of the difficulties encountered when applying the postulates of causality, it is accepted that not all criteria listed by authors need to be fulfilled.^{7,9-11,19} In case it is impossible to fulfill all the criteria, the evidence accumulated over time and the common sense of researchers will be important to identify the causal agent of a certain disease.^{9,13,19}

VIRAL METAGENOMICS

The term "metagenomics" indicates a joint analysis of microbial genomes in an environmental sample, not only from the genetic point of view but also in terms of function.¹⁸

The term "viral metagenomics" involves the detection of the genome of all viruses present in environmental samples (e.g., fresh water lake, reclaimed water),^{5,20} or biological samples (e.g., respiratory tract aspirates, human and animal feces)^{14,17,28} that could harbor a large diversity of viruses. This term is also used when the metagenomic approach is applied to identify the genome of a virus that can potentially cause a specific disease and/or a cytopathic effect in cell culture, when other common techniques failed to detect the virus.^{24,30}

The metagenomic approach includes several steps, as follows: the purification and concentration of the viral particles (or the viral nucleic acid if the virus is found in the latent form or integrated into the host genome), nucleic acid extraction, reverse transcription of RNA to cDNA, random amplification of genomic sequences, sequencing of nucleic acid fragments, and sequence

analysis using bioinformatics tools.^{2,12,27} Nucleic acid fragments can be sequenced using the Sanger method after molecular cloning or using next-generation sequencing platforms, which are more sensitive and generate a much larger number of sequences than molecular cloning using a bacterial host.^{2,12,16,27}

DISCOVERY OF VIRUSES THROUGH VIRAL METAGENOMICS

Although the metagenomic approach has been significantly contributing to the tremendous increase in the discovery of viruses,¹⁶ the number of novel associations between viruses and diseases has not been increasing in the same proportion. The causal association depends not only on detecting the presence of a virus in a sick person but also on conducting a complete investigation of the virus-disease association in order to comply with Koch's postulates or with the criteria of causation proposed later.

The use of the metagenomic approach in environmental samples has enabled the discovery of several novel genomic sequences potentially derived from viruses. However, the data obtained from these genomes are insufficient to identify the hosts and assess the pathogenic potential of the viruses. Cataloging these genomes into public databases is important, so that after further research, it will hopefully be possible to identify the viral hosts.

The need to identify the correct host and the potential pathogenicity of the virus is also imperative when a previously unknown virus is found in fecal samples or upper respiratory tract secretions. The presence of a virus in these samples during the acute phase of the disease does not necessarily make this agent responsible for the pathology. This can be the case when a virus shedded from the host for a prolonged period, e.g., enterovirus and bocavirus, is detected.²⁹ In addition, a virus detected in fecal samples or respiratory tract secretions may have been inhaled or ingested and may have passed through the lumen of the respiratory or digestive tract without replicating into that host.^{3,12}

The human bocavirus exemplifies the difficulty in evaluating the causal association between a newly discovered virus in the respiratory tract and the clinical manifestations. Bocaviruses were discovered in 2005 using a metagenomic approach in a pool of randomly selected samples of nasopharyngeal aspirates¹ and have been a topic of intense research since then. This research has indicated that the factors that hinder the establishment of a causal association between the virus and disease include the high prevalence of bocavirus infection, prolonged viral shedding by the host after infection, persistence of the viral DNA in the respiratory tract for several months, and high rate of coinfection. The studies conducted to date suggest

that bocaviruses are sometimes transient passengers and eventually pathogens of the respiratory tract.^{4,29}

Even when the metagenomic approach leads to the detection of a new virus in the cerebrospinal fluid (CSF), which is generally sterile, the disease cannot be associated with the virus before further investigation.²⁶ This hypothesis can be discussed on the basis of recent findings of Tan et al,²⁶ who found a new cyclovirus in CSF specimens of two patients with an acute infection of the central nervous system. After its identification, this virus was detected in 4.0% of 642 CSF samples from patients suspected of having an infection of the central nervous system; however, it was not detected in any of the 122 samples from patients with a noninfectious neurologic disease. The viral genome was also found in fecal samples of healthy children, which suggested food-borne or fecal-oral transmission route. In addition, it was found in animal feces, suggesting the existence of animal reservoirs for this virus. These authors affirmed that, considering the current knowledge, it is impossible to establish a causal association between this virus and the disease according to Koch's criteria or their adapted versions. For further assessment of this association, Tan et al²⁶ are attempting to isolate the virus in cell cultures or animal models and to detect a specific immune response. This is a justifiable caution because a virus found in CSF can be a coinfectious agent – which would play a secondary role in disease pathology and could increase disease severity or facilitate the entry of other pathogens –, or a latent virus that was reactivated because of an infectious/inflammatory process, or it can simply reflect the detection of a latent virus.^{6,22,23,26} This discussion is common in cases of detection in CSF of human herpesviruses that are disseminated through the hematogenous route, such as the Epstein-Barr virus and the human herpesvirus 6.^{6,22,23} The new cyclovirus could also be like the anelloviruses, which establish chronic active infections, may be devoid of pathogenicity (they can be components of the normal human microflora), and can be found in the central nervous system, blood, and several other body fluids.¹⁵

The viruses can also be detected by metagenomic approaches in chronic diseases; however, the causal association can be even more difficult,²⁷ as can be observed in a large amount of data on the Merkel cell polyomavirus (MCV). This virus was identified in human Merkel cell carcinoma (MCC) samples,⁸ and the investigation of the causal association between MCV and the disease began with the investigation of 10 MCC samples from different patients; of these, the viral genome was detected in eight samples. In 75.0% of these samples, viral DNA was integrated into the tumor genome in a clonal pattern, suggesting that the infection and integration process preceded the clonal expansion of the tumor cells. Control tissue samples

tested positive for the MCV genome in expressively lower percentage, and there was evidence that the number of copies of the viral genome in these positive samples was lower than that in the MCC samples.⁸ A high incidence of MCV among MCC cases has been confirmed in several countries, except in Australia.²¹ It is known that human infection with MCV occurs early, considering that the seroprevalence is 50.0% among individuals aged below 15 years.²¹ Studies with RNA interference and on the genetic changes of the viral genome integrated into MCC cells have indicated that MCV may contribute to the onset of MCC.²¹

When a new virus is detected in the blood of a patient with a disease that is probably infectious, only the identification of a new agent is also not proof of its causal relationship with the disease, as can be exemplified by the discovery of a new bunyavirus in China, which was named Henan fever (HNF) virus³⁰ or severe fever with thrombocytopenia syndrome virus (SFTSV).³¹ This virus was detected almost simultaneously by two research groups using a metagenomic approach in serum³⁰ and blood leucocytes;³¹ samples were obtained during the acute phase of the disease, which is characterized by fever, thrombocytopenia, and leukopenia. An extensive epidemiological, clinical, and laboratory investigation was conducted along with this discovery. In the laboratory, the evidence for this association included viral isolation, followed by visualization of the viral particles by electron microscopy,

detection of the viral genome, and positive serology in patient samples.^{30,31} The investigation by both research groups also included analysis of the control groups. When discussing the causal association between the HNF virus and the severe fever with thrombocytopenia syndrome, both the groups concluded that, although they could not completely fulfill Koch's postulates, there was a strong evidence indicative of this association.^{30,31} The fact that independent researchers confirmed these results^{25,30,31} corroborates this potential association, as stated in Lipkin's guidelines.¹³

FINAL CONSIDERATIONS

Viral metagenomics has been having a great impact on the discovery of new viruses because it enables the detection of all viral genomes present in a given sample independently from antisera tests, from previous knowledge of the viral genome (unlike other molecular biology techniques such as polymerase chain reaction, microarray, and *in situ* hybridization), and from cell culture isolation.^{2,16,27}

However, the causal association between a virus and a disease in humans and other animals still depends on a set of clinical, epidemiological, and laboratory investigations; on the use of strict criteria to associate these elements, such as those in Koch's postulates and their adapted versions; and on common sense during data analysis.^{9,10,12,13,16,19}

REFERENCES

- Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci U S A*. 2005;102(36):12891-6. DOI:10.1073/pnas.0504666102
- Ambrose HE, Clewley JP. Virus discovery by sequence-independent genome amplification. *Rev Med Virol*. 2006;16(6):365-83. DOI:10.1002/rmv.515
- Castrignano SB, Nagasse-Sugahara TK, Kisielius JJ, Ueda-Ito M, Brandão PE, Curti SP. Two novel circo-like viruses detected in human feces: complete genome sequencing and electron microscopy analysis. *Virus Res*. 2013;178(2):364-73. DOI: 10.1016/j.virusres.2013.09.018
- Debiaggi M, Canducci F, Ceresola ER, Clementi M. The role of infections and coinfections with newly identified and emerging respiratory viruses in children. *Virol J*. 2012;9:247. DOI:10.1186/1743-422X-9-247
- Djikeng A, Kuzmickas R, Anderson NG, Spiro DJ. Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One*. 2009;4(9):e7264. DOI:10.1371/journal.pone.0007264
- Doja A, Bitnum A, Jones ELF, Richardson S, Tellier R, Petric M, et al. Pediatric Epstein-Barr virus-associated encephalitis: 10-year review. *J Child Neurol*. 2006;21(5):384-91. DOI:10.2310/7010.2006.00114
- Evans AS. Causation and disease: the Henle-Koch postulates revisited. *Yale J Biol Med*. 1976;49(2):175-95.
- Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*. 2008;319(5866):1096-100. DOI:10.1126/science.1152586
- Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev*. 1996;9(1):18-33.
- Huebner RJ. Criteria for etiologic association of prevalent viruses with prevalent diseases; the virologist's dilemma. *Ann N Y Acad Sci*. 1957;67(8):430-8.
- Koch R. An address on bacteriological research. *Br Med J*. 1890;2(1546):380-3.
- Li L, Delwart E. From orphan virus to pathogen: the path to the clinical lab. *Curr Opin Virol*. 2011;1(4):282-8. DOI:10.1016/j.coviro.2011.07.006
- Lipkin WI. The changing face of pathogen discovery and surveillance. *Nat Rev Microbiol*. 2013;11(2):133-41. DOI:10.1038/nrmicro2949

14. Lysholm F, Wetterbom A, Lindau C, Darban H, Bjerkner A, Fahlander K, et al. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PLoS One*. 2012;7(2):e30875. DOI:10.1371/journal.pone.0030875
15. Maggi F, Bendinelli M. Human anelloviruses and the central nervous system. *Rev Med Virol*. 2010;20(6):392-407. DOI:10.1002/rmv.668
16. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol*. 2012;2(1):63-77. DOI:10.1016/j.coviro.2011.12.004
17. Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, Delwart EL. The fecal viral flora of wild rodents. *PLoS Pathog*. 2011;7(9):e1002218. DOI:10.1371/journal.ppat.1002218
18. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet*. 2004; 38:525-52. DOI:10.1146/annurev.genet.38.072902.091216
19. Rivers TM. Viruses and Koch's postulates. *J Bacteriol*. 1937;33(1):1-12.
20. Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M. Metagenomic analysis of viruses in reclaimed water. *Environ Microbiol*. 2009;11(11):2806-20. DOI:10.1111/j.1462-2920.2009.01964.x
21. Schrama D, Ugurel S, Becker JC. Merkel cell carcinoma: recent insights and new treatment options. *Curr Opin Oncol*. 2012;24(2):141-9. DOI:10.1097/CCO.0b013e32834fc9fe
22. Studahl M, Hagberg L, Rekabdar E, Bergström T. Herpesvirus DNA detection in cerebral spinal fluid: differences in clinical presentation between alpha-, beta-, and gamma-herpesviruses. *Scand J Infect Dis*. 2000;32(3):237-48. DOI:10.1080/0036554005016585
23. Sundén B, Larsson M, Falkeborn T, Paues J, Forsum U, Lindh M, et al. Real-time PCR detection of human herpesvirus 1-5 in patients lacking clinical signs of a viral CNS infection. *BMC Infect Dis*. 2011;11:220. DOI:10.1186/1471-2334-11-220
24. Svraka S, Rosario K, Duizer E, 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
25. Takahashi T, Maeda K, Suzuki T, Ishido A, Shigeoka T, Tominaga T, et al. The first identification and retrospective study of severe fever with thrombocytopenia syndrome in Japan. *J Infect Dis*. 2014;209(6):816-27. DOI:10.1093/infdis/jit603
26. Tan LV, Doorn HR, Nghia HDT, Chau TTH, Tu LTP, Vries M, et al. Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *MBio*. 2013;4(3):e00231-13. DOI:10.1128/mBio.00231-13
27. 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
28. Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, et al. 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
29. Williams JV. Déjà vu all over again: Koch's postulates and virology in the 21st century. *J Infect Dis*. 2010;201(11):1611-4. DOI:10.1086/652406
30. Xu B, Liu L, Huang X, Ma H, Zhang Y, Du Y, et al. Metagenomic analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan Province, China: discovery of a new bunyavirus. *PLoS Pathog*. 2011;7(11):e1002369. DOI:10.1371/journal.ppat.1002369
31. Yu XJ, Liang MF, Zhang SY, Liu Y, Li JD, Sun YL, et al. Fever with thrombocytopenia associated with a novel bunyavirus in China. *N Engl J Med*. 2011;364(16):1523-32. DOI:10.1056/NEJMoa1010095

The authors declare no conflict of interest.