《生物科学杂志(ISSN 2958-4035)》
Journal of Biological Sciences (ISSN 2958-4035)

ChengZhu Science™

Serial in November, 2022

ChengZhu Science Research Center Limited

(Hong Kong Registration NO. 80561228)

Journal article DOI is registered via JiangXi Province ChengZhu Environmental Engineering Company Ltd, with Crossref membership (DOI prefix 10.58473)

出版人： 刘焕 香江出版社有限公司

Publisher: Liu Huan, Xiangjiang Publishing Company Ltd.



Copyrights Statements

Copying and Transferring is Forbidden!

版权申明

禁止复制、转载！

All the intellectual property (mainly including the original academic knowledges and brand logo) are prohibited to copy or transfer into other publications or websites. To cite this article, only short quote is acceptable, but copying or transferring any substantial part of this article is NOT allowed (Defined in <Copyright Ordinance> in Hong Kong). The original academic knowledge is the substantial part of an article as academic journal. For learning purpose, it is allowed to read our website in online video class only. This journal is published by Hong Kong Publisher, and the copyrights is regulated and protected by <Copyright Ordinance> in Hong Kong, China. This PDF document is accessible to public only through Hong Kong domain websites (natural-foundation-science.org), and its printed version is the formally published journal. Without permission, it is NOT allowed to print, issue and sale.

所有形式知识产权（主要包括原创型学术知识和品牌标识）禁止复制、转载到其他出版物和网站。如果需要引用这篇论文，仅仅允许简短引述，但是禁止复制、转载这篇论文中任何实质性部分（香港《版权条例》中定义）。作为学术杂志，这篇论文中的原创型学术知识即为作品的实质性部分。仅仅允许以学习为目的在线视频课堂阅读本公司网站。本杂志由香港出版社出版，其版权受中国香港《版权条例》监管和保护。此PDF文档仅仅通过香港主机网站向公众公开 (natural-foundation-science.org)，并且其印刷版本杂志为正式出版物。未经许可，不得印刷、发行、销售。

Article 1-1. Classification of Virus by DNA Genetic Marker and Its Theory/DNA 遗传标记技术与病毒生物分类以及相关理论

Author: Liu Huan (1983 -), Master of Science (First Class Honours, 2009), The University of Auckland.

DOI: https://doi.org/10.58473/JBS0001

Formally published on 30/11/2022; Latest revised on 27/04/2026.

Abstract

This article presents the new experiment methods based on DNA genetic markers, which is the systematic method of classifying and identifying virus, and then the underlying theory of systematic classification is discussed.

Key words: Virus Classification, Genetic Marker, Virus Genotype, Virus Phenotype, Gene Recombination.

Methods

Step 1. The morphological markers of different virus samples are analyzed using transmission electron microscopy, and the criteria of morphological markers can be selected by referring to molecular cytogenetic karyotype analysis [1];

Step 2. Virus is classified by multivariate cluster analysis and genetic distance analysis on the basis of morphological markers, preliminarily leading to different families of virus [2];

Step 3. The optima sampling units of each virus family, which can well represent the genetic diversity of each virus family, is examined and determined as pointed out by Liu et al.,(2015) [2] for further classification based on DNA (or RNA) molecular marker (SSR or AFLP). The sampling units can be adjusted by changing the concentration of virus solution;

Step 4. Classification between preliminary virus families is further conducted on the basis of DNA (or RNA) molecular marker, analyzed by multivariate cluster analysis of UPGMA (unweighted pair group method with arithmetic averages) and genetic distance analysis [2].

Step 5. Gene recombination and gene mutation rate is analyzed by the inconsistency of classification between morphological markers and DNA (or RNA) molecular

marker.

There are three hypotheses examined by this research:
Hypothesis 1: the optima numbers of polymorphic SSR primers are examined and screened for each virus family, because we assume that the amount of polymorphic SSR primers, which are assessed on the basis of polymorphism information content (PIC), increases with the increase of total SSR primers selected from Gene Bank, but the increase rate is not constant. Consequently, the optimal number of polymorphic SSR primers is determined at the peak increase rate.

Hypothesis 2: the classification significantly differs between morphological markers of virus and DNA genetic markers.

Hypothesis 3: there are two kinds of multivariate cluster analysis and genetic distance analysis on the basis of SSR markers, resulting in two different classifications of virus families: firstly, the classification of virus families is conducted based on the Nei' genetic identity [3] (or Nei' genetic similarity[4]) calculated by the total SSR primers from Gene Bank; or then the classification of virus families is conducted based on the genetic identity calculated by the polymorphic SSR primers only. This research aims to examine which classification method leads to better correlation with the incidence of pathological characters recorded.

Discussion:

1.The total SSR primers selected from Gene Bank are the pairs of SSR primers which lead to clear PCR bands for at least one preliminary virus family in amplified process;

2.The virus samples should be collected in the same area at local scale, which facilitates the differentiation of local virus families due to the unique nature of virus ecosystem. There are two reasons to explain the virus ecosystem: Firstly, the local virus ecosystem is relatively isolated, due to 'the absence of gene communication' among virus ecosystem; Secondly, the SSR gene sequences which lead to clear PCR bands is considered to be highly conserved, so that its proportion of specific gene sequence information is relatively constant in its original virus ecosystem. Consequently, the virus samples should be collected in the same area at local scale. However, the amount of virus families, which result in the impacts on human health, is increasing due to gene recombination and mutation in self-reproduction process or invasive virus species from exotic ecosystem.

3.There is a novel matrix designed for genetic distance analysis based on PCR bands: Virus samples are collected routinely from the local virus ecosystem. Then preparation of DNA samples in one test: 12 uniform samples are abstracted from the same DNA water solution which has been evenly mixed, named as sample 1, sample 2, ..., sample12; In total 12 different SSR primers are selected in one test, named as

primer 1, primer 2,...., primer12, and each different SSR primer is injected into sample 1, sample 2, ..., sample 12 respectively for PCR amplified process; after PCR amplified process, each sample (12 in total) is electrophoresed separately in each pipe of electrophoresis instrument, and the PCR bands from different virus families, preliminarily drawn by morphological markers, would be clearly separated from each other in a electrophoresis pipe. Consequently, the distance between two PCR bands from two different virus families, which is measured in a pipe of electrophoretogram, represents the genetic distance between these two virus families per SSR primer (or locus). Then the multivariate cluster analysis of UPGMA (unweighted pair group method with arithmetic averages) is conducted on the basis of the average genetic distance between any two different virus families across SSR primers (or loci) which can lead to clear PCR bands in a pipe. If the electrophoresis pipe is the vertical one, then the PCR bands around the same horizontal lines would represent the same proportion of gene sequences information to the sum informations of the whole genome families examined in a experiment, which is comparable across different pipes in a matrix for UPGMA analysis. Once epidemic disease is broken out by the gene mutation or exotic virus species, this UPGMA analysis newly designed should accurately detect the variation as unbalanced status compared with original virus ecosystem, which suits for the routinely monitoring of local virus ecosystem.

There is an improved method presented for identification of virus families:
Step 1. The whole genome of a specific virus family, whose DNA (or RNA) molecular weight is measured in Lab [5], is cultivated for reproduction in Lab as standardized DNA molecule.

Step 2. After amplified process in PCR, the DNA fragment samples together with the cultivated genomes in step 1, are transferred into the electrophoretogram procedure.

Step 3. The standardized DNA (or RNA) molecule should be the molecules of the highest weight; Then the molecular weight of DNA fragments from the other virus families can be calculated per SSR correspondingly [5]. This improved method facilitates the identification of virus families, regardless of variation in virus ecosystem.

Step 4. Identification of virus family with gene mutation: the specific locus of genome, in which gene mutation occurs, is identified by DNA (or RNA) molecular markers (the heterozygous bands of a specific locus is the gene mutation bands, as compared to the homozygous bands of parental virus family without gene mutation). Please note: the heterozygous or homozygous bands here are just description of band morphology, rather than allelic gene.

Please note: the objects of dyeing procedure in step 1 is protein due to the protein ‘coat’ around virus DNA (or RNA) and the DNA (or RNA) molecules are the molecules with the highest weight in virus physiology, whereas the objects of dyeing

procedure in step 3 is nucleic acid molecule. SDS-PAGE for protein separation requires lower voltage than nucleic acid molecules (or isozyme separation), so that the DNA (or RNA) can take off their protein 'coat.' The weaker clearness of protein 'coat's bands, the higher accuracy of this test, which can be adjusted by gradual change of voltage. If SDS-PAGE for protein separation is conducted exactly, it is expected to reveal the morphology of virus DNA molecules by fluorescence in situ hybridization (FISH) technique, which can be analyzed by the criteria of molecular cytogenetic karyotype.

In this experiment, the gene mutation virus family is identified in the whole virus ecosystem, analyzed by both multivariate cluster method and two-paired comparison (between parental virus and gene mutation virus). It is hypothesized that the gene mutation virus family may show closer genetic distance to the other virus families, rather than its parental virus family, on the basis of the classification conducted by the DNA morphological markers of virus. However, the conclusion of virus classification is 'corrected' by further DNA (or RNA) sequencing markers (gene mutation virus family should show closer genetic distance to their parental virus family). This finding will further support the distortive bio-signal caused by gene mutation virus family, which is hardly identified by host cells discussed in another article of this journal [6].

In this article, it is further to propose that compared with the DNA sequences technology utilization only, the combination of both morphological markers and DNA sequences technology in this article results in different classification conclusions due to the relativity nature of statistics in Multivariate Classification Analysis, which is more reasonable for virus testing in terms of reflecting the whole DNA/genome information. This article further concludes that the fragment information of DNA sequences does not indicates the total pathological characters of virus, because the gene mutation does not only lead to the alteration of DNA sequences in a specific fragment, but also results in the significant changes in genome morphology which plays the key role in virus invasion process. Consequently, the virus testing based on the fragment DNA sequences only is not a reliable method to reveal the total gene mutation characters. This is especially important for COVID 19 testing.

My article has re-defined the gene recombination in virus as: "gene recombination does not limit to diploid sporophyte in the sexual reproduction process, but also applies on the microbes in the self-reproduction process. For example, virus may alter the spatial arrangement and combination information of existing DNA sequences (gene recombination) to adapt the changing environment in its self-reproduction process, rather than directly altering the existing DNA sequences by gene mutation. Gene mutation may be considered to be harder than gene recombination, which requires more generations' self-reproduction to occur. Consequently, the effectiveness of DNA sequencing technology may be limited in analyzing the genetic variation of a virus species in its self-reproduction process.[7]" Additionally, my another article has

re-defined the gene recombination in cell creatures as: "Gene recombination is considered as the spatial re-arrangement and re-combination between these conserved DNA sequence regions in cell division process. In the gene recombination process, cell mainly changes the DNA sequences in the binding regions, without altering the DNA sequences in the conserved and repeated DNA sequence regions during cell division process. In this way, cells pass on gene variation to offspring cells without significantly altering fundamental life functions expressed by the conserved DNA sequence regions (otherwise it may become the cancer cell).[8]" However, the gene recombination in virus would slightly differ from the cells, and it is proposed that gene recombination in virus should be re-defined as the spatial re-arrangement and re-combination between these conserved gene sequence regions in self-reproduction process. Virus creature, containing less fundamental life functions than cell creature, may contain less repetitive gene sequence regions, but it is hypothesized that virus tends to insert the repetitive gene sequences from the host cell's genome into virus genome molecule, facilitating the invasion against host cells.

In addition to Step1-Step5 above, there is my another article designing to identify the virus phenotype [9] as step 6, further classifying the virus strains.

References:

[1]. 刘焕, 张洪初与唐秋盛, 保护遗传学方法在生物多样性监测和评价领域的应用研究. 科技视界, 2014(8);
[2]. Liu, H, Ouyang T. L, Tian Chengqing (2015). Review of Evolutionary Ecology Study and Its Application on Biodiversity Monitoring and Assessment. Science & Technology Vision (6) 2015;
[3]. 陶玲与任珺, 进化生态学的数量研究方法, 2004, 中国林业出版社: 北京市;
[4]. Genuineness and Purity Verification of Potato Seed Tuber - SSR Molecular Marker (GB/T 28660-2012).
[5]. 朱广廉，杨中汉 SDS-聚丙烯酰胺凝胶电泳法测定蛋白质的分子量《植物生理学报》, 1982.
[6].Liu Huan, Gene Mutation, Pathogenesis and Gene Modification, Feb 2021, Journal of Environment and Health Science. https://doi.org/10.58473/JBS0007
[7].Liu Huan (2024). Original Review of Biodiversity Conservation Progress in China: The Theories and Cases in Practice. Journal of Environment and Health Science (ISSN 2314-1628). 2024(04). https://doi.org/10.58473/JEHS0012
[8].Liu Huan. (2021). Discussion on Gene Sequencing and Classification of DNA Genetic Markers. Journal of Environment and Health Science (ISSN 2314-1628), 2021(05). https://doi.org/10.58473/JBS0019
[9].Liu Huan. (2021). The Identification of Pathogen Phenotype for Biophysical Simulation. Journal of Environment and Health Science (ISSN 2314-1628), 2021(02). https://doi.org/10.58473/JBS0010