果 树 学 报 2019，36(5)：557-566
Journal of Fruit Science

DOI:10.13925/j.cnki.gsxb.20180490

香蕉条斑病毒*ORF Ⅰ*基因克隆及编码蛋白和同源物的生物信息学分析

黎维丽[1,2]，余乃通[2*]，李俊成[3]，王健华[1]，李宾宾[1,2]，刘志昕[2*]

（[1]海南大学热带农林学院，海口 570228；[2]中国热带农业科学院热带生物技术研究所•海南省微生物学重点实验室•农业农村部热带作物生物学与遗传资源利用重点实验室，海口 571101；[3]广东省农业科学院果树研究所•广东省热带亚热带果树研究重点实验室，广州 510640）

摘 要：【目的】探明杆状DNA病毒属病毒的*ORF Ⅰ*基因功能。**【方法】**本研究从香蕉条斑病毒云南分离株(BSV-YN)中克隆了BSV *ORF Ⅰ*基因，通过核苷酸和氨基酸序列分析和同源关系分析，确定了BSV-YN的分类地位；对其编码蛋白及同源物的理化性质、氨基酸序列组成、亲疏水性、跨膜区域、信号肽、亚细胞定位、磷酸化位点、同源结构域、蛋白二级和三级结构等进行预测和分析。**【结果】**BSV *ORF Ⅰ*基因片段大小为528 bp，编码175个氨基酸，通过核苷酸和氨基酸序列分析和同源关系分析，确定了BSV-YN属于BSGFV种。生物信息学分析表明，BSV-YN ORF Ⅰ及同源蛋白都是不稳定的、亲水性蛋白，以丝氨酸和苏氨酸磷酸化为主。这些蛋白二级结构主要由α-螺旋和无规则卷曲组成，无信号肽或跨膜区域。亚细胞定位预测表明，BSV-YN ORF Ⅰ及同源蛋白主要在细胞核分布，推测它们在病毒感染的宿主细胞核中发挥重要作用。**【结论】**推测BSV-YN ORF Ⅰ在病毒感染的宿主细胞核中参与病毒复制、转录等相关过程。本研究为进一步开展该蛋白的功能研究提供科学线索。

关键词：香蕉条斑病毒；*ORF Ⅰ*；同源蛋白；生物信息学分析

中图分类号：S668.1 **文献标志码：**A **文章编号：**1009-9980(2019)05-0557-10

Cloning of *Banana streak virus ORF Ⅰ* and Bioinformatic analysis of its protein and homologues

LI Weili[1,2], YU Naitong[2*], LI Juncheng[3], WANG Jianhua[1], LI Binbin[1,2], LIU Zhixin[2*]

([1]Institute of Tropical Agriculture and Forestry, Hainan University, Haikou 570228, Hainan, China; [2]Institute of Tropical Bioscience and Biotechnology, CATAS/Hainan Provincial Key Laboratory of Microbiology/Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture and Rural Affairs, Haikou 571101, Hainan, China; [3] Institute of Fruit Tree Research, Guangdong Academy of Agricultural Sciences/Guangdong Key Laboratory of Tropical and Subtropical Fruit Tree Research, Guangzhou 510640, Guangdong, China)

Abstract:【Objective】According to the report of the Tenth International Committee on Taxonomy of Viruses (ICTV) , genus *Badnavirus* has 46 species of viruses, including 9 distinctive species of *Banana streak virus* (BSV), which are *Banana streak GF virus* (BSGFV), *Banana streak UA virus* (BSUAV), *Banana streak IM virus* (BSIMV), *Banana streak MY virus* (BSMYV), *Banana streak OL virus* (BSOLV), *Banana streak UI virus* (BSUIV), *Banana streak UL virus* (BSULV), *Banana streak UM virus* (BSUML) and *Banana streak VN virus* (BSVNV). Badna viruses are widely distributed in tropical and temperate regions of the world and cause severe diseases in several important monocot and dicot crops, such as banana, yam, cocoa, citrus, black pepper and taro. Their genome size ranges from 7.2 to 9.2 kb with three to seven open reading frames (ORFs). However, the functions of many predicted ORFs including BSV *ORF Ⅰ* gene and its homologues have been kept unknown.【Methods】Banana leaves with black stripe, chlorotic stripe, spindle and eye-shaped symptoms of banana streak disease

收稿日期：2018-12-20 接受日期：2019-02-18

基金项目：广东省热带亚热带果树研究重点实验室开放基金(2017B030314113)；中国热带作物学会青年人才托举工程(CSTC-QN201704)；中国热带农业科学院基本科研业务费专项 (No.19CXTD-33)

作者简介：黎维丽，女，在读硕士研究生，研究方向为植物病毒学。Tel：0898-66890770，E-mail：1124047447@qq.com

*通信作者 Author for correspondence. E-mail：yunaitong@163.com；E-mail：liuzhixin@itbb.org.cn

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

(BSD) in Yunnan Province, China were collected in the year of 2008. The pathogen was tentatively named *Banana strike virus* Yunnan isolate (BSV-YN). The total DNA was extracted by using CTAB method and it was used as a DNA template for PCR amplification. The *ORF Ⅰ* gene of the BSV-YN was cloned and the classification relationship of BSV-YN in badna viruses was determined based on the *ORF Ⅰ* sequences analysis. The sequence similarity of *ORF Ⅰ* gene and its homologues genes were analyzed by NCBI blastn and BioEdit software, while the homology relationship of BSV-YN ORFⅠand its homologues proteins were conducted by MEGA6.0 with Maximum Likelihood method. In order to further investigate the characterization and function of ORFⅠand its homologues proteins, physicochemical properties, amino acid sequence composition, hydrophobicity, transmembrane structure, signal peptide, subcellular localization, phosphorylation site, structural functional domain, secondary and tertiary structures were predicted based on the online databases and softwares on bioinformatics websites.【Results】The BSV-YN *ORF Ⅰ* gene was 528 bp, encoding 175 amino acids with unknow function. NCBI blastn and BioEdit software analysis showed that BSV-YN *ORF Ⅰ* gene was 99.0% sequences similarity with BSGFV *ORF Ⅰ* gene (Genbank accession No. KJ013507.1), while its protein showed 100% identity with BSGFV ORFⅠ(Genbank accession No. AHM92949.1). Phylogenetic tree analysis indicated that BSV-YN ORFⅠhad the highest homology with the BSGFV ORFⅠ, followed by the BSUAV ORFⅠand *Commelina yellow mottle virus* (ComYMV) ORFⅠ.Therefore, BSV-YN belonged to the BSGFV species. ORFⅠs of BSV-YN, BSGFV, BSUAV and BSIMV had 175 to 177 amino acid residues with their relative molecular weights of 20.80 kDa. Furthermore, the isoelectric points (pIs) among 6.98 to 8.45, about 30 negative and 30 positive amino acid residues, the hydrophobicity among -0.741 to 0.606, and protein instability coefficients among 45.85 to 63.27 were also predicted. These results indicated that BSV-YN, BSGFV, BSUAV and BSIMV ORFⅠs were unstable, hydrophilic or water-soluble proteins. BSV-YN ORFⅠhad fourteen potential phosphorylation sites, including nine serine sites (Amino acid sites at 3, 14, 18, 32, 33, 34, 88, 103 and 126 in the peptide chain), four threonine sites (Amino acid sites at 45, 72, 148, 152 in the peptide chain) and one tyrosine (Amino acid site at 27 in the peptide chain). The results predicted that the serines and threonines of BSV-YN ORFⅠwere mainly phosphorylation sites. The secondary structure of BSV-YN ORFⅠand its homologous proteins consisted mostly of alpha helix (73.00% to 88.57%), followed by random coil (10%), extended strand (1.70% to 9.00%) and extended strand (less than 2%). Further analysis showed that the tertiary structure BSV-YN ORFⅠhad high similarity with BSGFV, BSIMV, SCBV and KTSV ORFⅠs, which were mainly consisted of alpha helix and random coil as well. Bioinformatic analysis indicated that BSV-YN ORFⅠand its homologous proteins did not have signal peptide region or transmembrane domain region. Subcellular localization prediction indicated that these proteins are mainly distributed in the cell nucleus. Therefore, BSV-YN ORFⅠand its homologous proteins were presumably involved in the viral infection process in the host cell nucleus.【Conclusion】The gene ORF of BSV-YN *ORF Ⅰ* was 528 bp long, encoding 175 amino acids with unknow function. Sequences similarity and homology relationship analyses of the BSV-YN *ORF Ⅰ* and homologues genes indicated BSV-YN belonged to the BSGFV species. Bioinformatic analysis showed that BSV-YN ORFⅠand homologous proteins were unstable, hydrophilic proteins, with serine and threonine phosphorylation. The secondary structure was mainly composed of α-helix and random coil in BSV-YN ORFⅠand homologous proteins, but no signal peptide or transmembrane domain of these proteins. Subcellular localization prediction indicated that these proteins were mainly distributed in the cell nucleus, presumably involved in the viral infection process. The proteins might play some key roles in the virus-infected cell nucleus, such as viral replication and transcription. This study would provides scientific clues for further research on the function of this protein.

Key words: *Banana streak virus*; *ORF Ⅰ*; Homologue protein; Bioinformatics analysis

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

香蕉作为人类重要的经济作物，在我国南方地区广泛种植。然而由香蕉条斑病毒（*Banana streak virus*，BSV）、香蕉束顶病毒（*Banana bunchy top virus*，BBTV）和黄瓜花叶病毒（*Cucumber mosaic virus*，CMV）等引起的病毒病严重威胁着香蕉产业的生产[1]。香蕉条斑病毒是侵染香蕉的重要病原物之一，是杆状拟反录病毒（Pararetroviruses），属于花椰菜花叶病毒科（Caulimoviridae）的杆状DNA病毒属（*Badnavirus*）[2-3]。BSV基因组为环状双链DNA，大小在7.5~8.0 kb [4]。BSV基因组转录产生大于一个环状全长基因组的病毒大RNA转录本，进入细胞核后通过自身编码的反转录酶（RNaseH）和宿主$tRNA_{met}$进行病毒基因组的逆转录复制[5-6]。另外，病毒大RNA转录本也可作为mRNA模板，在宿主细胞内编码病毒蛋白[5-6]。

根据第十次国际病毒分类委员会（ICTV）报告，杆状DNA病毒属含有46种病毒，引起重要单子叶和双子叶植物的病毒病害。由于BSV不同分离物的序列差异较大，结合BSV在世界各地的地理分布，ICTV把BSV分成9个种，分别为香蕉条斑GF病毒（*Banana streak GF virus*，BSGFV）、香蕉条斑UA病毒（*Banana streak UA virus*，BSUAV）、香蕉条斑IM病毒（*Banana streak IM virus*，BSIMV）、香蕉条斑MY病毒（*Banana streak MY virus*，BSMYV）、香蕉条斑OL病毒（*Banana streak OL virus*，BSOLV）、香蕉条斑UI病毒（*Banana streak UI virus*，BSUIV）、香蕉条斑UL病毒（*Banana streak UL virus*，BSULV）、香蕉条斑UM病毒（*Banana streak UM virus*，BSUML）和香蕉条斑VN病毒（*Banana streak VN virus*，BSVNV）[7]。香蕉条斑病毒基因组在一条链上有3个连续的ORFs，即*ORF Ⅰ*、*ORF Ⅱ*和*ORFⅢ*，分别编码3个蛋白，它们的功能还不清楚[8]。预测BSV *ORF Ⅱ*编码蛋白与核酸结合相关[9]。*ORFⅢ*编码一个多聚蛋白，推测经剪切后形成外壳蛋白（Coat protein，CP）、天冬氨酸蛋白酶（Asparticproteas，AP）、RNA酶H（RNase H，RH）和运动蛋白（Movement protein，MP）等结构和功能蛋白[10]。BSV *ORF Ⅰ*编码蛋白功能未知，但是与BSV同属的鸭跖草黄色斑驳病毒（Commelina yellow mottle virus，ComYMV）*ORF Ⅰ*与病毒粒子具有相互作用[11-13]，但两者的氨基酸序列相似性只有33.6%。

目前，90%以上的蛋白功能可通过生物信息学分析得到。通过获得基因的核苷酸序列和氨基酸序列，就能利用现有的开放数据库以及各类相关软件，预测和分析其结构和可能的生物学功能，为后续实验开展提供科学线索[14-16]。杆状DNA病毒*ORF Ⅰ*基因编码蛋白的功能鲜见报道。因此，笔者们从香蕉条斑病毒云南分离株克隆获得了*ORF Ⅰ*基因，并对其编码蛋白及同源物蛋白的特征和可能的功能开展生物信息学分析，为下一步开展该蛋白的功能研究提供借鉴。

1　材料和方法

1.1　材料及引物

香蕉条斑病毒云南分离株（BSV-YN）总DNA和载体pET32a由本实验室保存；大肠杆菌（Trans DH 5a）、高保真DNA聚合酶（PrimeSTAR HS DNA polymerase）、T4 DNA连接酶、限制性内切酶*Sal* Ⅰ和*Bam* HⅠ均购自大连宝生物有限公司；PCR产物凝胶回收试剂盒、质粒小量提取试剂盒、DL2000 DNA Marker均购自北京天根生化公司。

根据香蕉条斑病毒GF株系*ORF Ⅰ*全长编码框的核苷酸序列（GenBank登录号为KJ013507.1），设计扩增香蕉条斑病毒云南分离株*ORF Ⅰ*基因的一对特异性引物，*ORF Ⅰ*-F（5'-CGC GTC GAC ATG AAC TCC GAC CTC AAA GAG-3'，下划线为*Sal* Ⅰ酶切位点）和*ORF Ⅰ*-R（5'-CGC GGA TCC TCC AAG AAT TGT TTT TCT TAG ATG ATG-3'，下划线为*Bam* HⅠ酶切位点），由英潍捷基（广东）贸易有限公司合成。

1.2　方法

1.2.1　BSV-YN *ORF Ⅰ*基因的扩增　以BSV-YN总DNA为模版，*ORF Ⅰ*-F/R为引物进行PCR扩增。PCR反应体系如下：5×PrimeSTAR Buffer 10 μL，PrimeSTAR HS DNA polymerase 0.5 μL，dNTP Mixture 4 μL，正反向引物（5 μmol · L^{-1}）共2 μL，样品DNA模板2 μL，灭菌蒸馏水31.5 μL，总体积50 μL。PCR反应程序如下：98 ℃预变性3 min；98 ℃变性30 s，55 ℃退火30 s，72 ℃延伸30 s，共35个循环；最后72 ℃延伸10 min。取6 μL的PCR产物于1.5%琼脂糖凝胶电泳检测，并回收纯化。将回收纯化的PCR产物与原核表达载体pET32a分别经*Bam* HⅠ和*Sal* Ⅰ双酶切和回收。再利用T4 DNA连接酶将PCR产物连接至pET32a载体中，转化大肠杆菌

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

Trans DH 5a感受态细胞，经测序确定BSV-YN *ORF Ⅰ*的核苷酸序列。测序由英潍捷基（广东）贸易有限公司完成。

1.2.2 *ORF Ⅰ*及同源物序列分析及进化树构建 BSV-YN *ORF Ⅰ*测序结果经ChromasPro软件进行序列编辑和拼接后，将其核苷酸和编码氨基酸序列经NCBI中的Blastn和Blastp分别进行搜索比对，下载相似性较高的10种同属病毒*ORF Ⅰ*s的核苷酸和编码氨基酸序列（表1），再经BioEdit和GeneDoc软件对其及同源物的氨基酸序列分别进行联配和序列相似性分析。利用MEGA 6.0软件的最大似然法（Maximum Likelihood Method）构建系统发育树，分析BSV-YN ORFⅠ和其他10种同源ORFⅠs的亲缘关系。

表1 BSV-YN和10种同属病毒信息

Table 1 BSV-YN and 10 homology viruses information

病毒中文名 Virus (Chinese)	病毒英文命名 Virus (English)	Genbank登录号 Genbank accession No.
香蕉条斑病毒云南株	*Banana streak virus* YN *isolate* (BSV-YN)	–
香蕉条斑GF病毒	*Banana streak GF virus* (BSGFV)	AHM92949.1
香蕉条纹UA病毒	*Banana streak UA virus* (BSUAV)	YP_004442822.1
香蕉条纹IM病毒	*Banana streak IM virus* (BSIMV)	YP_004442837.1
甘蔗杆状病毒	*Sugarcane bacilliform virus* (SCBV)	AJF46482.1
鸭跖草黄色斑驳病毒	*Commelina yellow mottle virus* (ComYMV)	NP_039818.1
美人蕉黄斑驳病毒	*Canna yellow mottle-associated virus* (CaYMV)	YP_009259696.1
菠萝杆状CO病毒	*Pineapple bacilliform CO virus* (PBV)	YP_003987463.1
苏铁叶坏死病毒	*Cycad leaf necrosis virus* (CLNV)	YP_002117529.1
嵌条叶条纹病毒	*Enset leaf streak virus* (ELSV)	AXP07688.1
伽蓝菜顶端斑点病毒	*Kalanchoe top-spotting virus* (KTSV)	NP_777315.1

1.2.3 生物信息学分析 由于BSV-YN ORFⅠ和其他10种ORFⅠs同源蛋白的功能尚不清楚，因此，本研究利用生物信息学分析它们的基本特征和预测可能的蛋白功能。利用在线网页ExPASy Proteomic的ProtParam tool程序分析BSV-YN ORFⅠ蛋白和10种同源物的分子质量、等电点和氨基酸组成等理化性质；利用在线网页EMBL的SMART程序和NCBI的Conserved Domain程序预测蛋白质的同源结构域；利用ProtScale在线工具分析蛋白质的亲疏水性；利用SOPMA在线工具预测蛋白的α-螺旋、β-折叠及其他二级结构；利用在线软件TMHMM 2.0 Server预测蛋白质的跨膜结构域区域；利用SignalP 4.1 Server在线工具对蛋白质的信号肽区域进行预测；利用PSORTII在线工具预测蛋白质可能的亚细胞定位结果；利用NetPhos 3.1 Server在线工具分析蛋白质潜在的磷酸化位点；利用SWISS-MODEL软件在线预测蛋白质三级结构（表2）。

表2 生物信息学研究的数据库和软件网站

Table 2 Online databases and softwares used for bioinformatics analysis

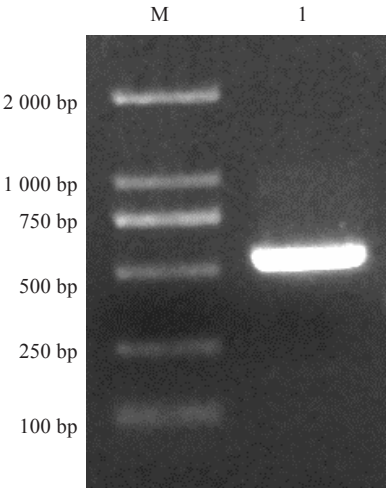
工具 Tool	网页 Website
ProtParam tool	http://web.expasy.org/protparam/
ProtScale tool	http://web.expasy.org/ProtScale /
SOPMA	http://npsa-prabi.ibcp.fr/cgi-bin/scpred_sopma.pl/
TMHMM 2.0 Server	http://www.cbs.dtu.dk/servers/TMHMM/
SignalP 4.1 Server	http://www.cbs.dtu.dk/servers/SignalP/
SMART	http//smart.embl-heidelberg.de/
PSORTII	http://psort.hgc.jp/form.html
NetPhos 3.1 Server	http://www.cbs.dtu.dk/services/NetPhos/
SWISS-MODEL	http: //www.swissmodel.expasy.org/
Conserved Domain	http: //blast.ncbi.nlm.nih.gov/
BLASTn/BLASTp	http: //blast.ncbi.nlm.nih.gov/

2 结果与分析

2.1 BSV-YN *ORF Ⅰ*基因的扩增

PCR结果如图1所示，扩增PCR产物在500 bp处得到一条特异的DNA条带（Lane 1），与预期BSV-YN *ORF Ⅰ*片段大小一致。测序结果和NCBI blastn分析结果表明，该基因片段长度为528 bp，与BSGFV的*ORF Ⅰ*基因序列（Genbank accession No. KJ013507.1）高度相似（99.0%），仅有两个核苷酸的差异（A102G，A474G）。推测其编码175个氨基酸的蛋白，与已收录的BSGFV ORFⅠ蛋白（Genbank accession No. AHM92949.1）氨基酸序列一致，二者

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



M. Tran 2 000 bp DNA marker; 1. BSV-YN *ORF Ⅰ*的 PCR 产物。

M. Tran 2 000 bp DNA marker; 1. PCR products of *ORF Ⅰ* from BSV-YN isolate.

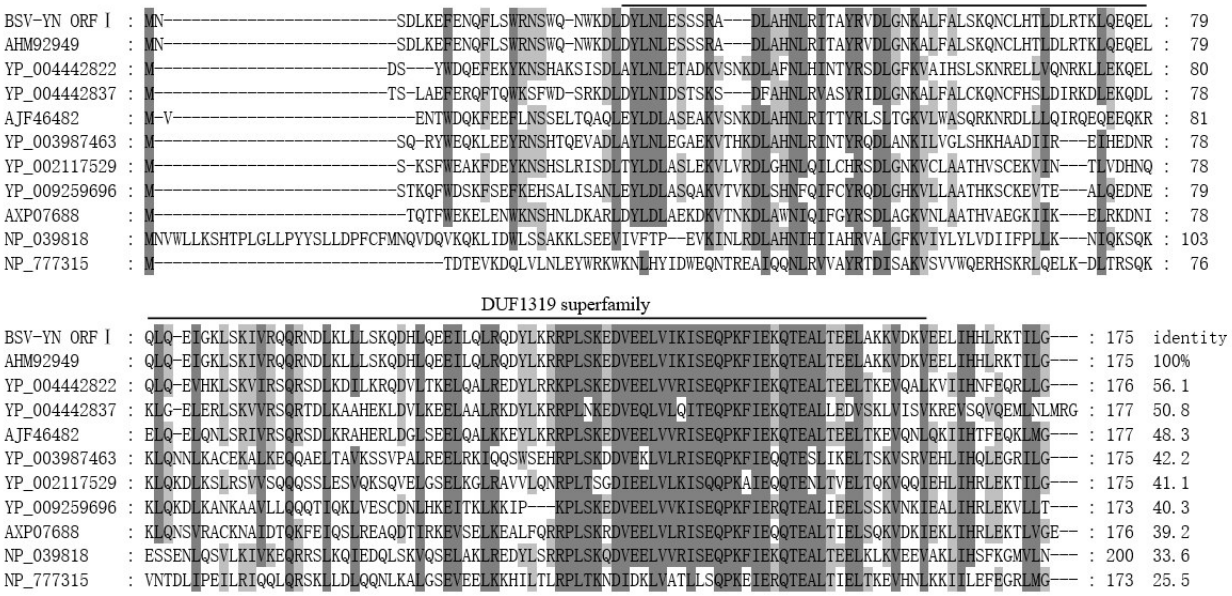
图1　BSV-YN *ORF Ⅰ*的PCR扩增

Fig. 1 PCR amplification of BSV-YN *ORF Ⅰ*

序列相似性为100%。

2.2　ORFⅠ氨基酸序列比对与同源分析

氨基酸序列分析表明，本研究的BSV-YN ORFⅠ与所选的10种同源蛋白的序列相似性为25.5%~100%（图2）；其中与香蕉条斑病毒不同种的BSGFV、BSUAV和BSIMV的ORFⅠs序列相似性较高，分别为100%、56.1%和50.8%。同源结构域预测表明，这11个蛋白均含有一个保守的DUF1319超家族结构域，该结构域主要为杆状DNA病毒属病毒所特有。利用MEGA 6.0软件的最大似然法构建系统发育树，分析这11个蛋白的亲缘关系。结果表明，BSV-YN与香蕉条斑病毒的BSGFV和BSUAV种亲缘关系较近，其中与BSGFV亲缘关系最接近。有意思的是，BSV-YN虽然与BSIMV的序列相似性较高，但是其同源性较ComYMV、CaYMV、PBV、CLNV和ELSV低。说明BSV病毒在不同地理位置具有不同



该蛋白含有功能未知的保守结构域（DUF1319 超家族），该家庭成员主要为杆状 DNA 病毒属病毒所特有。

The protein contains a conserved domain (DUF1319 superfamily) with unknown function. The family members seem to be restricted to badnaviruses.

图2　BSV-YN ORFⅠ和同源物氨基酸的多序列比对

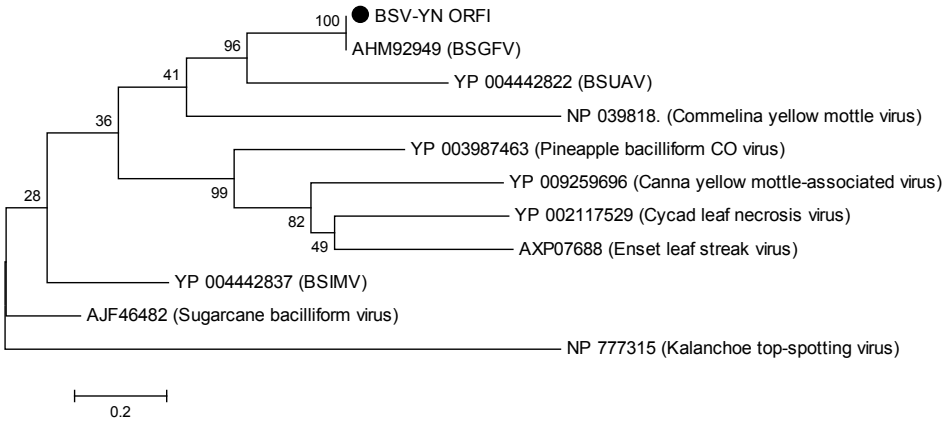
Fig. 2 Multiple amino acid sequences alignment of BSV-YN ORFⅠand homology proteins

程度的进化途径，形成特有的分支（图3）。因此，本研究分离的香蕉条斑病毒云南分离株（BSV-YN）属于BSGFV种。

2.3　蛋白质一级结构分析

氨基酸序列组成及理化性质分析表明（表3），BSV-YN ORFⅠ与所选的10种同源蛋白ORFⅠs的氨基酸残基数为173~200个，相对分子质量为19.83~23.33 kDa，等电点（pI）为6.04~9.05，负电荷残基数（Asp + Glu）为22~32，正电荷残基数（Arg + Lys）为22~32，脂肪族氨基酸系数为96.99~112.08。不稳定性系数为43.74~65.51，亲疏水性预测范围为-0.882~-0.169，表明这11个蛋白均为不稳定的（>

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net



使用 MEGA6 构建进化树[17]；基于 JTT 矩阵模型，使用最大似然法推断同源关系[18]。

Evolutionary analyses were conducted in MEGA6 [17]. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [18].

图 3 利用最大似然法构建 BSV-YN ORF Ⅰ和同源物的系统发育分析

Fig. 3 Molecular Phylogenetic analysis of BSV-YN ORF Ⅰ and homologues by Maximum Likelihood method

表 3 BSV-YN ORF Ⅰ蛋白及同源物的一级结构分析

Table 3 Primary structure analysis of BSV-YN ORF Ⅰ and homology proteins

病毒 Virus	长度 Length/aa	分子质量 MW/kDa	等电点 pI	天冬氨酸 Asp (-)	谷氨酸 Glu (-)	精氨酸 Arg (+)	赖氨酸 Lys (+)	不稳定性系数 Instability coefficient	脂肪族氨基酸系数 Aliphatic-aa coefficient	亲疏水性 Hydrophobicity
BSV-YN	175	20.80	6.98	11	18	11	18	63.27	105.31	-0.741
BSGFV	175	20.80	6.98	11	18	11	18	63.27	105.31	-0.741
BSUAV	177	20.80	8.45	14	16	13	19	45.85	96.38	-0.606
BSIMV	176	20.80	7.88	11	19	11	20	55.15	102.44	-0.710
SCBV	177	21.03	6.04	7	25	14	16	65.51	94.75	-0.882
ComYMV	200	23.33	9.01	8	18	8	22	59.57	119.30	-0.169
CaYMV	173	19.90	7.02	8	19	3	24	54.46	102.02	-0.453
PBV	175	20.44	7.14	7	21	12	16	51.93	97.54	-0.802
CLNV	175	19.83	7.06	8	14	7	15	58.87	110.74	-0.362
ELSV	176	20.60	7.79	11	20	11	21	43.74	96.99	-0.717
KTSV	173	20.53	9.05	9	17	12	17	53.75	112.08	-0.605

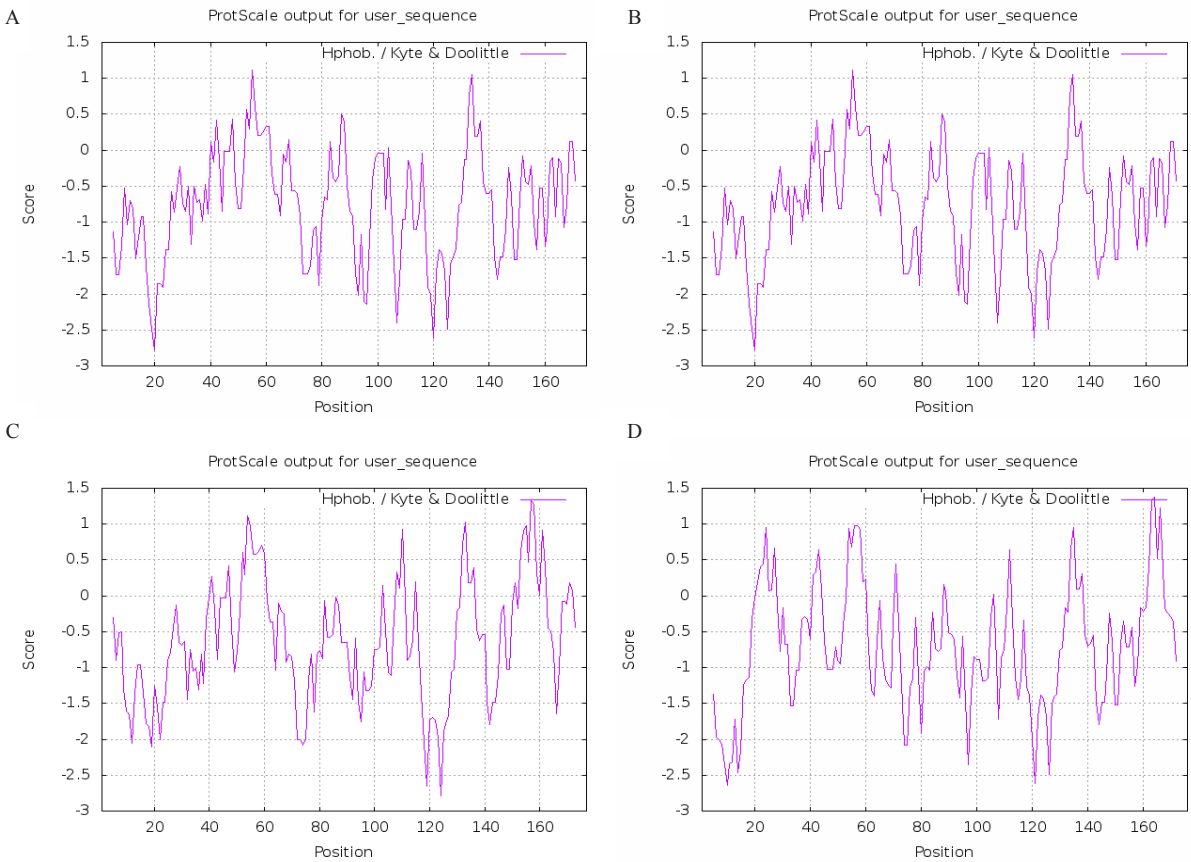
40)、亲水性或水溶性蛋白(负值)。进一步分析BSV-YN、BSGFV、BSUAV和BSIMV等4个BSV病毒种的ORF Ⅰ s，结果表明它们的氨基酸残基数为175~177个，相对分子质量均为20.80 kDa，等电点(pI)为6.98~8.45，所带负电氨基酸残基与所带正电氨基酸残基均在30个左右，亲疏水性为-0.741~-0.606，其蛋白的不稳定系数为45.85~63.27，属于不稳定的、亲水性或水溶性蛋白，与ProtScale在线工具分析的亲疏水性结果一致(图4)。

由表4可知，这11个蛋白质氨基酸序列均由20种氨基酸组成，且不同病毒ORF Ⅰ蛋白氨基酸所占比例相差不大，其中亮氨酸(Leu)、谷氨酸(Glu)和赖氨酸(Lys)的含量均较多，而半胱氨酸(Cys)的含量最少，推测与病毒蛋白自身的特性相关。

2.4 蛋白质二、三级结构预测和分析

蛋白质二级结构主要包括α-螺旋、β-折叠、无规则卷曲和延伸链。通过二级结构，蛋白质进一步形成更高级的三级结构，从而发挥蛋白的活性和功能。利用SOPMA在线分析表明，BSV-YN ORF Ⅰ及同源物蛋白二级结构主要元件是α-螺旋，含量在73%~88.57%，无规则卷曲含量在10%左右，延伸链在1.70%~9.00%，β-折叠含量大部分在2%以下(表5)。通过使用在线工具SWISS-MODEL对该蛋白质的三级结构进行预测，结果如图5所示，BSV-YN、BSGFV、BSIMV、SCBV和KTSV ORF Ⅰ s的三级结构相似性较高；而PBV、CLNV和ELSV的ORF Ⅰ s

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



横坐标为氨基酸的位置，纵坐标为亲疏水性得分（疏水性和亲水性分别用正值和负值表示）。

The abscissa is the position of the amino acid, while the ordinate is the hydrophobicity score. Hydrophobicity (positive value) and hydrophilicity (negative values) are showed.

图 4　BSV-YN (A)、BSGFV (B)、BSUAV (C)和 BSIMV (D)的 ORF Ⅰs 亲疏水性预测

Fig. 4 Hydrophobicity prediction of BSV-YN (A), BSGFV (B), BSUAV (C) and BSIMV (D) ORF Ⅰs

表 4　BSV-YN ORF Ⅰ及同源物氨基酸组成及含量分析

Table 4 Amino acid composition and content analysis of BSV-YN ORF Ⅰ and homology proteins　　%

AA	BSV-YN	BSGFV	BSUAV	BSIMV	SCBV	ComYMV	CaYMV	PBV	CLNV	ELSV	KTS
Ala	4.0	4.0	5.6	4.5	4.5	3.5	6.9	6.9	3.4	6.8	3.5
Arg	6.3	6.3	7.3	6.2	7.9	4.0	1.7	6.9	4.0	6.2	6.9
Asn	5.1	5.1	3.4	4.5	4.0	3.5	3.5	4.6	4.0	5.1	4.6
Asp	6.3	6.3	7.9	6.2	4.0	4.0	4.6	4.0	4.6	6.2	5.2
Cys	0.6	0.6	1.1	0.0	0.0	0.5	1.7	0.6	1.7	0.6	0.0
Gln	8.6	8.6	6.2	6.8	9.6	6.5	6.9	7.4	9.1	6.2	8.1
Glu	10.3	10.3	9.0	10.8	14.1	9.0	11.0	12.0	8.0	11.4	9.8
Gly	1.7	1.7	1.7	1.1	1.7	1.5	0.6	2.3	3.4	2.3	1.7
His	2.9	2.9	1.7	2.8	1.7	2.5	3.5	5.1	4.0	1.7	2.3
Ile	5.1	5.1	3.4	5.1	4.0	7.5	5.8	6.3	5.1	7.4	6.9
Leu	18.3	18.3	15.3	15.3	15.8	15.5	13.9	13.1	16.0	11.9	16.2
Lys	10.3	10.3	10.7	11.4	9.0	11.0	13.9	9.1	8.6	11.9	9.8
Met	0.6	0.6	1.7	0.6	1.1	1.5	0.6	0.6	0.6	0.6	1.2
Phe	2.3	2.3	4.0	2.8	2.3	3.5	3.5	0.6	1.1	2.8	0.6
Pro	1.1	1.1	1.1	1.1	1.1	3.5	1.7	1.7	1.1	1.1	1.7
Ser	6.3	6.3	7.3	7.4	6.8	8.0	7.5	6.9	10.3	4.5	4.0
Thr	3.4	3.4	3.4	3.4	5.1	2.0	4.6	3.4	4.6	5.1	6.9
Trp	1.7	1.7	1.1	0.6	1.1	1.0	0.6	1.1	0.6	1.7	2.3
Tyr	1.7	1.7	1.7	2.8	1.7	2.5	1.2	2.3	1.1	1.1	1.7
Val	3.4	3.4	6.2	6.2	4.5	9.0	6.4	5.1	8.6	5.1	6.4

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表5 BSV-YN ORF Ⅰ蛋白及同源物的二级结构预测

Table 5 Secondary structure prediction of BSV-YN ORF Ⅰ and homology proteins %

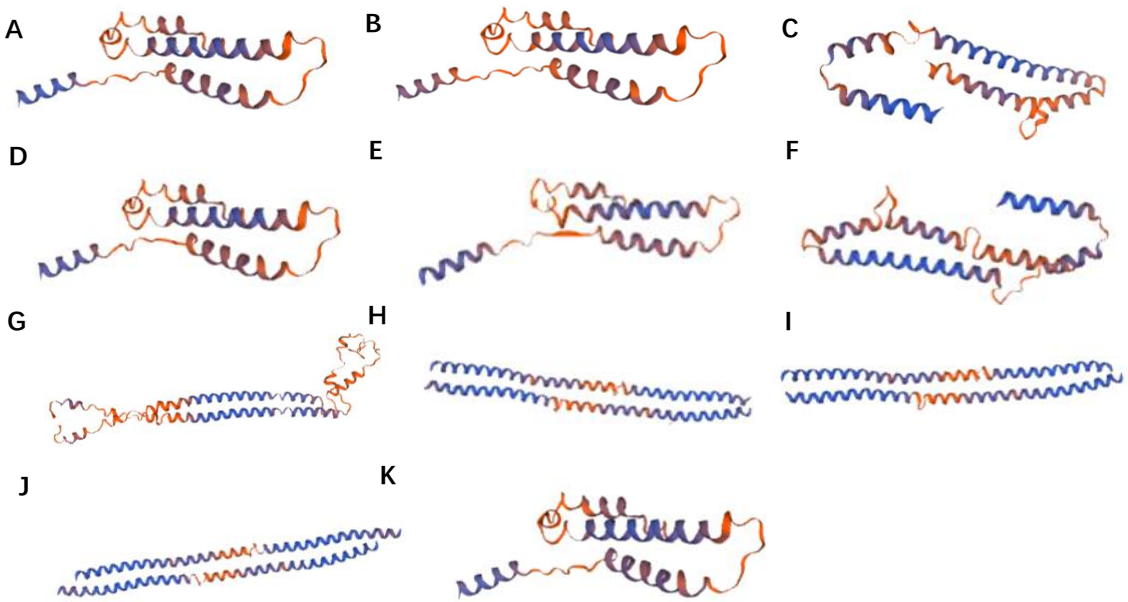
病毒 Virus	α-螺旋 Alpha helix	β-折叠 Beta turn	延伸链 Extended strand	无规则卷曲 Random coil
BSV-YN	83.43	0.57	3.43	12.57
BSGFV	83.43	0.57	3.43	12.57
BSUAV	84.18	1.13	4.52	10.17
BSIMV	84.66	0.57	3.41	11.36
SCBV	83.62	0.56	2.82	12.99
ComYMV	73.00	2.50	9.00	15.50
CaYMV	86.71	0.00	4.62	8.67
PBV	88.57	0.57	3.43	7.43
CLNV	82.86	0.57	2.86	13.71
ELSV	87.50	0.00	1.70	10.80
KTSV	83.24	1.73	5.78	9.25

三级结构相似。这些蛋白的三级结构主要由α螺旋和无规则卷曲这两种二级结构构成。

2.5 蛋白质磷酸化修饰预测

磷酸化是蛋白质最常见的修饰之一，在基因表达、信号传递和代谢等重要生命过程中起到关键调节作用。利用NetPhos 3.1 Server在线预测BSV-YN ORF Ⅰ及同源物蛋白潜在的磷酸化位点。结果显示(阈值0.5)，这11个蛋白磷酸化位点数在13~18个，其中潜在的丝氨酸(Ser)磷酸化位点最多，为6~13个，苏氨酸(Thr)为2~7个，酪氨酸(Tyr)最少，为0~2个(表6)。

BSV-YN ORF Ⅰ存在14个潜在的磷酸化位点，包括9个丝氨酸(Ser，位于肽链的第3、14、18、32、33、34、88、103、126位氨基酸)、4个苏氨酸(Thr，位于肽链的第45、72、148、152位氨基酸)和1个酪氨酸(Tyr，位于肽链的第27位氨基酸)。BSGFV、BSUAV和BSIMV ORF Ⅰs结果与BSV-YN ORF Ⅰ相似。因此，推测BSV-YN ORF Ⅰ及10个同源蛋白的磷酸化主要以丝氨酸和苏氨酸磷酸化为主。



BSV-YN (A)，BSGFV (B)，BSUAV (C)，BSIMV (D)，SCBV (E)，ComYMV (F)，CaYMV (G)，PBV (H)，CLNV (I)，ELSV (J)，KTSV (K) ORF Ⅰs.

图5 BSV-YN ORF Ⅰ及同源物的三级结构预测

Fig. 5 Tertiary structure prediction of BSV-YN ORF Ⅰ and homology proteins

表6 BSV-YN ORF Ⅰ及同源物的磷酸化位点分析

Table 6 Phosphorylation site analysis of BSV-YN ORF Ⅰ and homology proteins

病毒 Virus	丝氨酸 Ser	苏氨酸 Thr	酪氨酸 Tyr
BSV-YN	9	4	1
BSGFV	9	4	1
BSUAV	10	4	1
BSIMV	10	3	3
SCBV	7	7	1
ComYMV	13	3	0
CaYMV	8	6	1
PBV	10	6	2
CLNV	12	2	2
ELSV	6	6	2
KTSV	7	7	1

2.6 蛋白质的跨膜区域、信号肽区域和亚细胞定位预测

利用TMHMM 2.0在线软件对BSV-YN ORF Ⅰ及同源蛋白的跨膜区域预测，结果显示，只有Co-

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

mYMV ORF Ⅰ可能存在一个跨膜区域（73~92 aa），而其他同源蛋白均不存在跨膜区域。利用SignalP 4.1 Server在线软件预测这些蛋白的信号肽区域，结果都显示BSV-YN ORF Ⅰ及同源蛋白都不具有信号肽区域。

蛋白的亚细胞定位往往与蛋白质功能密切相关，在细胞不同位置分布就能发挥着不同的生物学功能，因此，分析蛋白质在细胞中的分布位置对未知蛋白功能研究具有重要借鉴意义。利用PSORTII在线预测BSV-YN ORF Ⅰ及这10同源蛋白的亚细胞定位，结果显示（表7），这些同源蛋白主要在细胞核分布，而ComYMV和KTSV的ORF Ⅰs主要分别在内质网和细胞质分布。结合上述蛋白质亲疏水性分析结果，由此推测，BSV-YN、BSGFV、BSUAV和BSIMV的ORF Ⅰs在病毒感染过程中的宿主细胞核中发挥作用，可能参与病毒复制、转录等相关过程。

表7 BSV-YN ORF Ⅰ蛋白及同源物亚细胞定位预测分析

Table 7 Subcellular localization prediction of BSV-YN ORF Ⅰ and homology proteins

%

病毒 Virus	细胞核 Nuclear	细胞质 Cytoplasm	线粒体 Mito	内质网 ER	细胞骨架 Cytoskeleton	细胞质膜 PM	高尔基体 Golgi	过氧化物酶体 Peroxisoma	液泡膜 VM
BSV-YN	60.9	21.7	13.0	4.3	-	-	-	-	-
BSGFV	60.9	21.7	13.0	4.3	-	-	-	-	-
BSUAV	60.9	8.7	4.3	-	8.7	4.3	8.7	4.3	-
BSIMV	52.2	21.7	-	-	13.0	4.3	4.3	4.3	-
SCBV	78.3	4.3	17.4	-	-	-	-	-	-
ComYMV	11.1	11.1	11.1	55.6	-	-	11.1	-	-
CaYMV	60.9	-	39.1	-	-	-	-	-	-
PBV	43.5	17.4	30.4	-	4.3	-	-	-	4.3
CLNV	60.9	8.7	30.4	-	-	-	-	-	-
ELSV	52.2	8.7	13.0	-	8.7	4.3	8.7	4.3	-
KTSV	17.4	60.9	17.4	4.3	-	-	-	-	-

3 讨　论

目前，香蕉条斑病毒（BSV）引起的香蕉条斑病毒病（*Banana streak disease*，BSD）在东南亚和非洲地区的香蕉和芭蕉等种植地流行和危害，是制约这两个区域香蕉产业发展的重要因素之一[1]。另外，由于BSV序列能整合到香蕉宿主基因组中（endogenous BSV，eBSV），在特定的时间能激活并感染宿主，因此BSV是一种潜在的、爆发性强的重要病毒[19]。研究表明，能整合到香蕉宿主基因组的BSV种有BSOLV、BSGFV、BSIMV和BSMYV，能引起感染的有BSOLV、BSGFV和BSIMV[20-22]。笔者们分离得到的BSV云南分离株属于BSGFV，既能整合到宿主基因组中，在适当的条件下又能引起感染，产生病毒粒子。因此，研究和阐明BSOLV、BSGFV和BSIMV这类香蕉条斑病毒的致病机理，对探索有效预防BSD和生产无eBSV香蕉苗具有重要的科学意义和应用价值。

BSV三个ORFs的编码蛋白功能尚不明确，且BSV *ORF Ⅰ*基因功能未知。因此，笔者从感染BSV的云南分离株克隆了*ORF Ⅰ*基因，基因片段大小为528 bp，编码175个氨基酸。通过核苷酸和氨基酸序列分析和同源关系分析，笔者们确定了来自中国云南分离株的香蕉条斑病毒属于BSGFV种。有意思的是，BSV-YN虽然与BSIMV的ORF Ⅰ序列相似性较高，但是其同源性较ComYMV、CaYMV、PBV、CLNV和ELSV低，说明BSV病毒在不同地理位置具有不同程度的进化途径，形成特有的分支。

另外，对BSV-YN *ORF Ⅰ*编码蛋白及杆状病毒属的同源蛋白的基本特征和可能的功能进行生物信息学分析，为下一步开展该蛋白的功能研究提供借鉴。BSV-YN ORF Ⅰ及同源蛋白质氨基酸序列均由20种氨基酸组成，且不同病毒ORF Ⅰ蛋白氨基酸所占比例相差不大，其中亮氨酸（Leu）、谷氨酸（Glu）和赖氨酸（Lys）的含量均较多，而半胱氨酸（Cys）的含量最少，推测与病毒蛋白自身的特性相关。半胱氨酸是形成二硫键的桥梁，从而影响蛋白的高级结构和生物功能。而BSV-YN ORF Ⅰ及同源蛋白质氨基酸中半胱氨酸含量很少，因此，推测这些同源蛋白通过很少的半胱氨酸维持或形成二硫键，从而发挥生物学功能。

跨膜域预测结果表明，除ComYMV的ORF Ⅰ存在一个跨膜区域，其它同源蛋白均未预测到跨膜区域。亚细胞定位分析显示，ComYMV的ORF Ⅰ主

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

要在内质网上分布。进一步说明ComYMV的ORF Ⅰ与其他BSV等同源蛋白存在差异。这可能与ComYMV的宿主鸭跖草与其他同源病毒宿主的进化关系较远相关。虽然不同病毒种可能归属于同一个病毒属，但是由于它们宿主之间的亲缘关系较远，在进化过程中病毒为了适应宿主环境，不同病毒蛋白就很可能发挥不同的功能作用[23]。

参考文献 References：

[1] KUMAR P L，SELVAJAN R，ISKRA-CARUANA M L. Biology，etiology，and control of virus diseases of banana and plantain [J]. Advances in Virus Research，2015，91(1): 229-269.

[2] KRUPOVIC M，BLOMBERG J，COFFIN J M. *Ortervirales*: A new viral order unifying five families of reverse-transcribing viruses[J]. Journal of Virology，2018，92(12): e00515-18.

[3] LEFKOWITZ E J，DEMPSEY D M，HENDRICKSON R C. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV)[J]. Nucleic Acids Research，2018，46(D1): D708-D717.

[4] BHAT AI，HOHN T，SELVARAJAN R. Badnaviruses: The current global scenario[J]. Viruses，2016，8(6): 177.

[5] HARPER G，HULL R. Cloning and sequence analysis of *Banana streak virus* DNA[J]. Virus Genes，1998，17(3): 271-278.

[6] HARPER G，HULL R，LOCKHART B. Viral sequences integrated into plant genomes [J]. Trends in Plant Science，2000，5(9): 362-365.

[7] ADAMS M J，LEFKOWITZ E J，AMP K. Changes to taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2017)[J]. Archives of Virology，2017, 162(8): 2505-2538.

[8] GEIJSKES R J，BRAITHWAITE K S，DALE J L. Sequence analysis of an Australian isolate of *sugarcane bacilliform badnavirus*[J]. Archives of Virology，2002，147(12): 2393-2404.

[9] JACQUOT E，HAGEN LS，JACQUEMOND M. The open reading frame 2 product of *Cacao swollen shoot badnavirus* is a nucleic acid-binding protein[J]. Virology，1996，225(1): 191-195.

[10] ZHUANG Jun，WANG Jianhua，ZHANG Xin，LIU Zhixin. Molecular characterization of *Banana streak virus* isolate from Musa Acuminata in China[J]. Virologica Sinica，2011，26 (6): 393-402.

[11] CHENG C P，LOCKHART B E L，OLSZEWSKI N E. The ORFI and II proteins of *Commelina Yellow Mottle Virus* are virion-associated[J]. Virology，1996，223(2): 263-271.

[12] MEDBERRY S L，LOCKHART B E，OLSZEWSKI N E. Properties of *Commelina yellow mottle virus's* complete DNA sequence，genomic discontinuities and transcript suggest that it is a pararetrovirus[J]. Nucleic Acids Research，1990，18(18): 5505-5513.

[13] 孙生仁，张慧丽，吴小斌 . 甘蔗杆状病毒及其启动子功能的研究进展[J]. 植物生理学报，2018，54(6): 943-950.
SUN Shengren，ZHANG Huili，WU Xiaobin. Research progress of *sugarcane baculovirus* and its promoter function[J]. Acta Physiologica Sinica，2018，54 (6): 943-950.

[14] 徐建华，朱家勇 . 生物信息学在蛋白质结构与功能预测中的应用[J]. 医学分子生物学杂志，2005，2(3): 227-232.
XU Jianhua，ZHU Jiayong. Application of bioinformatics in protein structure and function prediction[J]. Journal of Medical Molecular Biology，2005，2(3): 227-232.

[15] 贺光 . 生物信息学在蛋白质研究中的应用[J]. 国际遗传学杂志，2002，25(3): 156-158.
HE Guang. Application of bioinformatics in protein research[J]. International Journal of Genetics，2002，25(3): 156-158.

[16] 邓浪，沈兵琪，王连春 . ‘红阳’猕猴桃全基因组 AP2/EREBP 转录因子生物信息学分析[J]. 果树学报，2017，34(7): 790-805.
DEND Lang，SHEN Bingqi，WANG Lianchun. Bioinformatics analysis of the whole genome AP2/EREBP transcription factors of ‘Hongyang’ kiwifruit[J]. Journal of Fruit Science，2017，34 (7): 790-805.

[17] TAMURA K，STECHER G，PETERSON D，FILIPSKI A，KUMAR S. MEGA6: Molecular evolutionary genetics analysis version 6.0[J]. Molecular Biology and Evolution，2013，30(12): 2725-2729.

[18] JONES D T，TAYLOR W R，THORNTON J M. The rapid generation of mutation data matrices from protein sequences [J].Computer Applications in the Biosciences，1992，8(3): 275-282.

[19] DHONT A，DENOEUD F，AURY J M，BAURENS F C，CARREEL F，GARSMEUR O，NOEL B，BOCS S，DROC G，ROUARD M. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants[J]. Nature，2012，488 (7410): 213-217.

[20] STAGINNUS C，ISKRA-CARUANA M L，LOCKHART B. Suggestions for a nomenclature of endogenous pararetroviral sequences in plants[J]. Archives of Virology，2009，154(7): 1189-1193.

[21] ISKRA-CARUANA M L，BAURENS F C，GAYRAL P. A four-partner plant-virus interaction: enemies can also come from within[J]. Molecular Plant-Microbe Interactions，2010，23(11): 1394-1402.

[22] ISKRA-CARUANA M L，CHABANNES M，DUROY P O. A possible scenario for the evolution of *Banana streak virus* in banana[J].Virus Research，2014，186: 155-162.

[23] 张明霞，李琦，韩占英 . 汉坦病毒的进化及与宿主动物的关系[J]. 现代预防医学，2006，33(7): 1115-1117.
ZHANG Mingxia，LI Qi，HAN Zhanying. Evolution of Hantavirus and its relationship with host animals[J]. Modern Preventive Medicine，2006，33(7): 1115-1117.

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net