

## 三叶青烟草花叶病毒 lncRNA 测序鉴定、蛋白表达及其序列分析

洪森荣<sup>1,2,3,4</sup>, 徐燕萍<sup>1</sup>, 郭依瑶<sup>1</sup>, 章成<sup>1</sup>

1. 上饶师范学院生命科学学院, 江西 上饶 334001

2. 上饶农业技术创新研究院, 江西 上饶 334001

3. 上饶市药食同源植物资源保护与利用重点实验室, 江西 上饶 334001

4. 上饶市薯芋类作物种质保存与利用重点实验室, 江西 上饶 334001

**摘要:** **目的** 通过 lncRNA 测序鉴定三叶青 *Tetrastigma hemsleyanum* 烟草花叶病毒蛋白基因。**方法** 利用大肠杆菌重组技术表达三叶青烟草花叶病毒蛋白, 并采用生物信息学方法进行序列分析。**结果** 三叶青烟草花叶病毒蛋白基因包括复制酶、RNA 聚合酶、运动蛋白、带电蛋白和外壳蛋白等基因; 三叶青烟草花叶病毒 RNA 聚合酶、运动蛋白 (MP)、带电蛋白和外壳蛋白 (CP) 可实现大肠杆菌重组表达, 表明三叶青烟草花叶病毒的 lncRNA 测序结果准确; 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白基因 cDNA 总长度分别为 480、807、1 425、123 bp, 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白分别由 159、268、474、40 个氨基酸组成; 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶为亲水性蛋白质, 带电蛋白为疏水性蛋白质; 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶二级结构由  $\alpha$  螺旋、 $\beta$ -片层、无规则卷曲构成, 带电蛋白二级结构由  $\beta$ -片层和无规则卷曲构成, 三级结构均为单体, 分别定位于叶绿体、叶绿体、细胞质和线粒体中。三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白与烟草花叶病毒亲缘关系较近, 同源性达 98% 以上。**结论** 三叶青烟草花叶病毒具有烟草花叶病毒的结构特征, 氨基酸序列及核酸序列与同源物种相似度高, 在进化上高度保守。

**关键词:** 三叶青; 烟草花叶病毒; lncRNA 测序鉴定; 蛋白表达; 序列分析

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2024)20-7072-13

DOI: 10.7501/j.issn.0253-2670.2024.20.022

## lncRNA sequencing identification, protein expression and sequence analysis of tobacco mosaic virus of *Tetrastigma hemsleyanum*

HONG Senrong<sup>1,2,3,4</sup>, XU Yanping<sup>1</sup>, GUO Yiyao<sup>1</sup>, ZHANG Cheng<sup>1</sup>

1. College of Life Sciences, Shangrao Normal University, Shangrao 334001, China

2. Shangrao Agricultural Technology Innovation Research Institute, Shangrao 334001, China

3. Key Laboratory of protection and utilization of medicinal and edible plant resources in Shangrao City, Shangrao 334001, China

4. Shangrao Key Laboratory of germplasm conservation and utilization of potato and taro crops, Shangrao 334001, China

**Abstract: Objective** To identify the protein gene of tobaccos mosaic virus in *Tetrastigma hemsleyanum* by sequencing of lncRNA. **Methods** Proteins was expressed by *Escherichia coli* recombinant technology, and the sequence was analyzed by bioinformatics methods. **Results** The results showed that the protein genes of tobaccos mosaic virus in *T. hemsleyanum* included replicase, RNA polymerase, motor protein, charged protein and coat protein genes; The recombinant expression of its RNA polymerase, movement protein (MP), charged protein and coat protein (CP) could be realized in *E. coli*, which indicated that the sequencing result of lncRNA was accurate; The total cDNA length of CP, MP, RNA polymerase and charged protein genes of tobaccos mosaic virus in *T. hemsleyanum* were 480, 807, 1 425 and 123 bp, respectively. The CP, MP, RNA polymerase and charged protein were composed of 159, 268, 474 and 40 amino acids, respectively; The CP, MP, RNA polymerase was hydrophilic protein, and the charged protein was hydrophobic protein; The secondary structure of CP, MP, RNA polymerase was composed of alphahelix, extendedstran and randomcoil and the secondary structure of charged

收稿日期: 2024-04-22

基金项目: 国家自然科学基金资助项目 (31860084); 国家自然科学基金资助项目 (31960079); 国家自然科学基金资助项目 (32060092); 江西省现代农业产业技术体系建设专项 (JXARS-13-赣东站)

作者简介: 洪森荣 (1974—), 教授, 主要从事植物生物技术研究工作。Tel: (0793)8153721 E-mail: hongsenrong@163.com

protein was composed of extended strand and random coil. Their tertiary structures were all monomer, which were located in chloroplast, chloroplast, cytoplasm and mitochondria respectively. The CP, MP, RNA polymerase and charged protein of *T. hemsleyanum* were closely related to tobacco mosaic virus, with more than 98% homology. **Conclusion** TMV of *T. hemsleyanum* had the structural characteristics of tobacco mosaic virus, whose amino acid sequence and nucleic acid sequence were highly similar to those of homologous species and highly conservative in evolution.

**Key words:** *Tetrastigma hemsleyanum* Diels et Gilg; tobacco mosaic virus; lncRNA sequencing identification; protein expression; sequence analysis

三叶青 *Tetrastigma hemsleyanum* Diels et Gilg, 别名金丝吊葫芦, 为葡萄科崖爬藤属, 是我国特有珍稀濒危药用植物<sup>[1]</sup>。中医学认为三叶青全株均可入药, 可清热解毒、消肿止痛、化痰散结, 内服对小儿高热惊厥、肝炎、扁桃体炎等疗效显著; 外用可治毒蛇咬伤、跌打损伤等<sup>[2]</sup>。怀玉山产的三叶青, 是江西省上饶市玉山县特产, 也全国农产品地理标志<sup>[3]</sup>。怀玉山产的三叶青在生产上一般采用扦插无性繁殖, 极易感染烟草花叶病毒, 导致种性退化、产量下降、品质变劣<sup>[4]</sup>。因此, 对三叶青烟草花叶病毒进行 lncRNA 测序鉴定、蛋白表达及其序列分析, 对三叶青脱毒苗的培育和应用具有重要意义。关于三叶青烟草花叶病毒研究的报道较少, 如林国卫等<sup>[5]</sup>对三叶青的病毒进行了 RT-PCR 鉴定; 尹明华等<sup>[4,6-7]</sup>对三叶青的烟草病毒增殖蛋白 2B isoform X1 (TOM2B isoform X1)、烟草病毒增殖蛋白 3 (TOM3) 和烟草病毒增殖蛋白 2A (TOM2A) 进行了基因克隆和功能分析; 洪森荣等<sup>[3]</sup>对三叶青的烟草病毒增殖蛋白 1 (TOM1) 进行了基因克隆、亚细胞定位和组织表达分析。烟草花叶病毒 (tobacco mosaic virus, TMV) 属帚状病毒科 (Vigaviridae) 烟草花叶病毒属 *Tobamovirus*, 病毒粒子为杆状, 异常稳定, 侵染性能保持数十年<sup>[8]</sup>。lncRNA (long non-coding RNA) 是一类长度超过 200 bp, 没有开放阅读框的内源性非编码 RNA, lncRNA 建库时去除 rRNA, 可降低 rRNA 序列的干扰, 有利于病毒序列的分析<sup>[9]</sup>。目前, 对三叶青的研究主要集中在化学成分<sup>[10]</sup>、提取工艺<sup>[11]</sup>、药理作用<sup>[12]</sup>、栽培育苗<sup>[13]</sup>、种质资源<sup>[14]</sup>、光合作用<sup>[15]</sup>、组织培养<sup>[16]</sup>等方面, 关于三叶青烟草花叶病毒外壳蛋白 (coat protein, CP)、运动蛋白 (movement protein, MP)、RNA 聚合酶 (RNA polymerase)、带电蛋白 (charged protein) 基因方面的研究未见报道。本实验通过 lncRNA 测序鉴定三叶青烟草花叶病毒蛋白基因, 利用大肠杆菌重组技术表达三叶青

烟草花叶病毒蛋白, 并采用生物信息学方法对其进行序列分析, 为揭示三叶青脱毒苗的培育和鉴定提供理论依据。

## 1 材料与仪器

### 1.1 材料

三叶青栽培种“怀玉 2 号”试管苗由上饶市三叶青保育与利用技术创新中心提供, 采自怀玉山道地产区。该试管苗由上饶市红日农业开发有限公司陈荣华高级农艺师鉴定为三叶青 *T. hemsleyanum* Diels et Gilg。

### 1.2 仪器

NovaSeq 6000 型基因测序仪 (美国 Illumina 公司)、HH-4 型四孔数显恒温水浴锅 (南北科仪北京科技有限公司)、Allegra V-15R 型冷冻型高速台式离心机 (美国贝克曼库尔特有限公司)、MJQ70-I 型不锈钢四方立式蒸汽灭菌器 (上海寰熙医疗器械有限公司)、SW-CJ-2D 型垂直水平全钢超净工作台 (南京晓晓仪器设备有限公司)、XY-ZDT-82N 型台式恒温摇床 (上海昕仪仪器仪表有限公司)、Sonics VCX150PB 型超声波破碎仪 (美国 Sonics 公司)、TDEC06386FA 型超低温卧式冰箱 (赛默飞世尔科技公司)、AKTA Pure 型蛋白纯化仪 (美国 Cytiva 公司)、DYCZ-24DH 型 SDS-PAGE 电泳仪 (北京六一生物科技有限公司)、Qubit 4.0 型核酸蛋白定量荧光计 (赛默飞世尔科技公司)。

## 2 方法

### 2.1 三叶青烟草花叶病毒蛋白基因的确定

对三叶青栽培种“怀玉 2 号”试管苗的进行 lncRNA 测序。

**2.1.1 RNA 提取** 从三叶青栽培种“怀玉 2 号”试管苗植株上各采取 100 mg 左右幼嫩叶组织, 混合后提取小 RNA。

**2.1.2 RNA 质检** 完成 RNA 抽提后, 利用 1% 琼脂糖凝胶电泳检测 RNA 的质量; RNA 片段化, 400~500 bp, 反转录。

**2.1.3 文库构建** cDNA 片段末端补平, 加 A, 加测序接头; 琼脂糖凝胶电泳进行片段筛选, 将目标区域的片段回收; 通过 PCR 扩增筛选未加上接头的片段; PCR 产物纯化, 质检。

**2.1.4 桥式 PCR** DNA 片段的一端与芯片上引物碱基互补, 固定在芯片上; 另一端随机与附近的另一种引物互补, 也被固定住, 形成“桥(bridge)”;

PCR 扩增, 产生 DNA 簇; 线性化成为单链。

**2.1.5 Illumina HiSeq 4000 测序** 加入改造过的 DNA 聚合酶和带有 4 种荧光标记的 dNTP, 每次循环只掺入单种碱基; 用激光扫描反应板表面, 读取每条模板序列第 1 轮反应所聚合上去的核苷酸种类; 将“荧光基团”和“终止基团”化学切割, 恢复 3'端粘性, 继续聚合第 2 个核苷酸; 统计每轮收集到的荧光信号结果, 获知模板 DNA 片段的序列。利用 SOAPdenovo v2.04 (<http://soap.genomics.org.cn/>) 拼接软件对质控后的全部 clean data 进行 denovo 初步组装, 将组装得到的全部 contig 与病毒数据库进行比对, 筛选得到来自病毒的 contig; 然后利用 MITObim v1.6 通过迭代比对, 将测序的所有 clean reads mapping 到这些 contig 上进行延伸, 通过 gap close 获得病毒的全基因组序列。利用 DOGMA 软件 (<http://dogma.cccb.utexas.edu/>) 对基因组中包含的 gene 进行预测。将预测基因的蛋白序列与 Nr 数据库进行 blastp 比对 (BLAST 2.2.28+), 初步判断三叶青烟草花叶病毒外壳蛋白 (CP)、运动蛋白 (MP)、RNA 聚合酶、带电蛋白基因。

## 2.2 三叶青烟草花叶病毒蛋白的大肠杆菌重组表达

**2.2.1 感受态转化及阳性克隆筛选** 从超低温冰箱中取出 BL21 (DE3) 感受态细胞, 置于冰上融化; 加入质粒 (5  $\mu$ g), 轻轻吹吸充分混匀, 冰上放置 30 min; 水浴锅 42  $^{\circ}$ C 热击 90 s, 冰上放置 1 min; 加入 800  $\mu$ L 预热 LB 液体培养基, 37  $^{\circ}$ C, 158 r/min 培养 50 min; 6 000 r/min 离心 4 min, 去部分上清 (800  $\mu$ L), 剩余菌液混匀后涂至氨苄抗性平板上; 倒置平板, 37  $^{\circ}$ C 培养 12~16 h 可出现菌落 (阳性克隆)。

**2.2.2 阳性克隆小量表达与鉴定** 挑选含重组质粒的单菌落至 3 mL LB 液体培养基 (氨苄抗性) 中, 37  $^{\circ}$ C 培养过夜后 -20  $^{\circ}$ C 保种; 分别挑选含重组质粒的单菌落至 3 mL LB 液体培养基 (氨苄抗性) 中, 37  $^{\circ}$ C 震荡培养至 600 nm 吸光度 (A) 约 0.6; 取部分菌液作为对照组, 余下菌液加入 IPTG 诱导剂 (终

浓度 0.1 mmol/L), 16  $^{\circ}$ C 震荡培养 12 h; 分别取 2 组菌液 0.15 mL, 12 000 $\times$ g 离心 2 min, 菌体沉淀以 40  $\mu$ L 1 $\times$ loading buffer 重悬裂解, SDS-PAGE 检测。

**2.2.3 蛋白大量表达及破菌检测** 取保存于 -20  $^{\circ}$ C 的菌种 100  $\mu$ L 接种于 100 mL LB 液体培养基 (氨苄抗性) 中震荡培养过夜; 取 100 mL 菌液接种于 2 000 mL LB 液体培养基中, 37  $^{\circ}$ C 扩大培养至吸光度 ( $A_{600}$ ) 约 0.6, 降低培养温度到 30  $^{\circ}$ C; 加入 IPTG 诱导剂至终浓度 0.1 mmol/L, 16  $^{\circ}$ C 继续震荡培养 12 h; 8 000 r/min 离心 3 min 收集菌体, 重悬于 50 mL 预冷 NTA-0 缓冲液中, 冰浴 30 min; 超声破碎菌体, 参数设置为功率 200 W、工作 3 s、暂停 4 s、99 个循环; 4  $^{\circ}$ C、16 000 r/min 离心 50 min, 收集上清以及沉淀; 取少量上清及沉淀进行 SDS-PAGE 检测, 剩余上清及沉淀置于 4  $^{\circ}$ C 备用。

## 2.3 三叶青烟草花叶病毒蛋白基因的生物信息学分析

使用 BioEdit 软件翻译基因序列为氨基酸序列, 用 ProtParam 预测酶的理化性质, 用 ProtScale 预测酶的疏/亲水性。使用 GOR1 软件在线预测酶的二级结构。使用 SWISS-MOLD 在线预测酶的三级结构。采用 WoLFPSort 在线预测基因的表达部位。通过软件 DNAMAN 和 Bioedit 进行氨基酸序列比对, 利用 MEGA 5.0 软件进行系统进化树的构建。

## 3 结果与分析

### 3.1 原始测序数据碱基错误率分布统计

测序错误率会随着测序序列长度的增加而升高, 这是由测序过程中化学试剂的消耗导致的, 并且此现象为 Illumina 高通量测序平台的共有特征。由图 1 可知, 该文库碱基质量良好, 可用于后续分析。

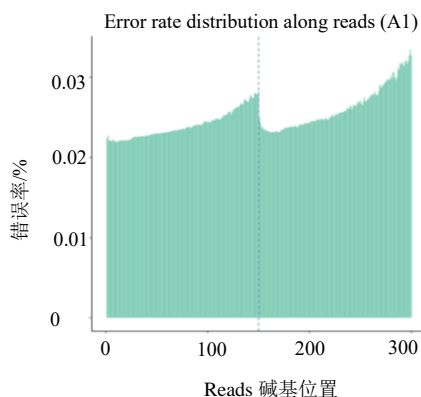


图 1 样品碱基错误率分布图

Fig. 1 Distribution of base error rate of samples

### 3.2 测序数据统计

对经过质量剪切前后的数据分别进行测序总碱基数、文库平均插入长度、 $Q_{20}$ 、 $Q_{30}$ 、平均测序深度统计, clean reads 为 92 237 536,  $Q_{20}$  和  $Q_{30}$  分别为 97.98 和 94.18, 均超过 85%, 表明建立的文库准确率很高。

### 3.3 基因组组装

基因组组装结果显示, 三叶青病毒种类是 1 个, 该病毒的总长度为 6 395 bp, 其 GC 含量为 43.35%。

### 3.4 Gene 查找

利用 DOGMA 软件对基因组中包含的 gene 进行预测。Gene 预测结果可知, 基因数量为 6 个, 基因总长度为 11 037 bp, 基因平均长度为 1 839.5, 基因间区 GC 含量为 43.35%。

### 3.5 基因功能注释

基因注释结果如表 1 和图 2 所示。三叶青烟草花叶病毒蛋白基因包括复制酶、RNA 聚合酶、运动蛋白、带电蛋白和外壳蛋白等。

表 1 基因功能注释结果

Table 1 Gene function annotation results

基因名称	链	起始	终止	基因长度/bp	蛋白质长度/aa	注释结果
ORF1	+	69	4 919	4 851	1 617	复制酶
ORF2	+	69	3 419	3 351	1 117	复制酶
ORF3	+	3 495	4 919	1 425	475	RNA 聚合酶
ORF4	+	4 903	5 709	807	269	运动蛋白
ORF5	+	5 666	5 788	123	41	带电蛋白
ORF6	+	5 712	6 191	480	160	外壳蛋白

Accession:123456  
Leng:Short

■ GC Skew+  
■ GC Skew-  
■ GC Content  
■ CDS

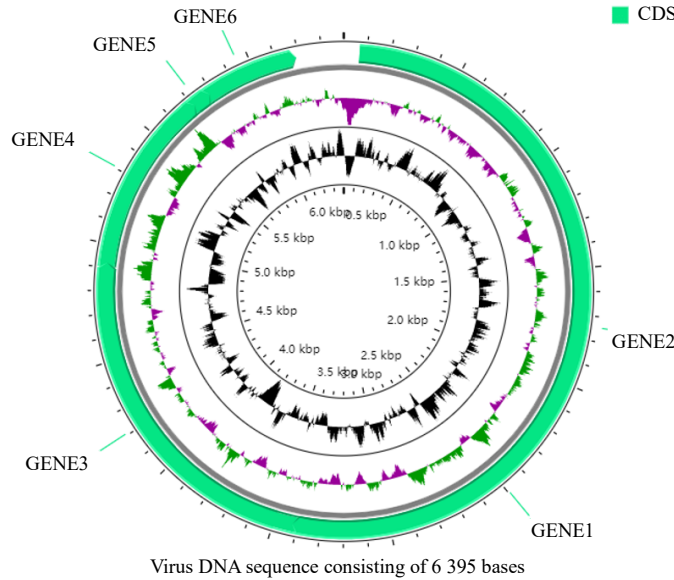


图 2 三叶青烟草花叶病毒基因组圈图

Fig. 2 Genome circle of tobacco mosaic virus in *T. hemsleyanum*

### 3.6 三叶青烟草花叶病毒蛋白的大肠杆菌重组表达

转化 BL21 (DE3) 感受态细胞, 挑选 AMP 抗性平板上长出的阳性斑进行扩大培养与表达纯化, 结果见图 3。ORF1 和 ORF2 的蛋白太大 (大于 100 000), 超过了大肠杆菌异源表达的能

力, 未能成功表达出目的蛋白。其他的 4 个目的蛋白 (ORF3、ORF4、ORF5 和 ORF6) 均诱导纯化后蛋白大小相符, 蛋白表达纯化成功。这说明, 三叶青烟草花叶病毒蛋白基因的 lncRNA 测序结果是准确的。

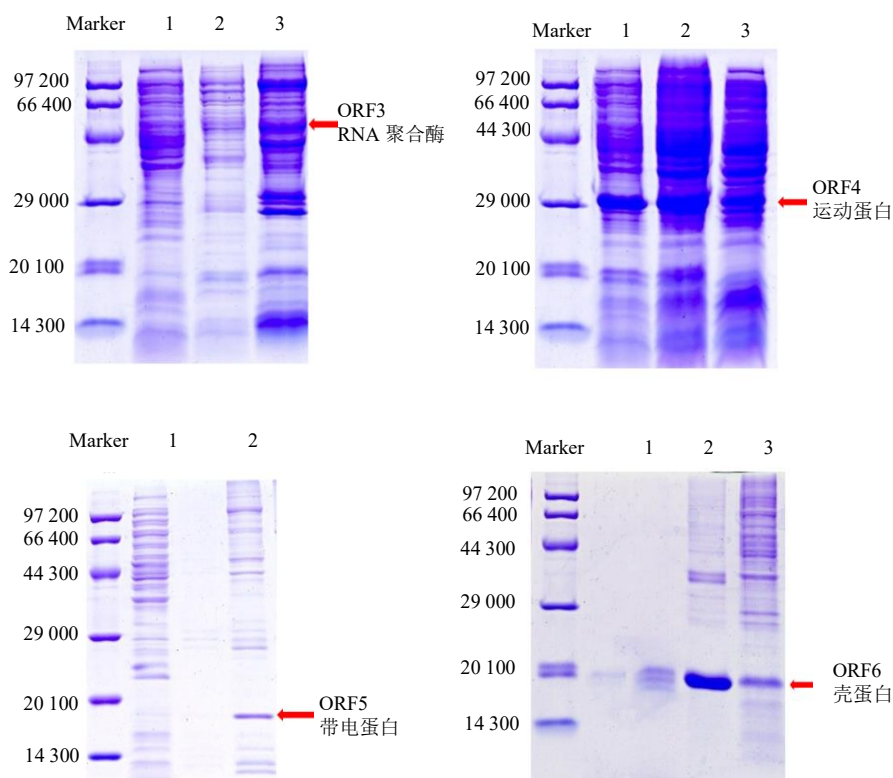


图 3 三叶青烟草花叶病毒 ORF3、ORF4、ORF5 和 ORF6 蛋白表达

Fig. 3 Protein expression of ORF3, ORF4, ORF5 and ORF6 of tobacco mosaic virus in *T. hemisleyanum*

### 3.7 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白基因 cDNA 序列

三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白基因 cDNA 总长度分别为 480、807、1 425、123 bp(图 4), G+C 含量为 42.71%、40.40%、40.98% 和 40.65%。

### 3.8 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白氨基酸序列

三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白氨基酸序列见图 5。三叶青烟草花叶病毒 CP 由 159 个氨基酸组成, 相对分子质量 17 636.73, 等电点 5.09。三叶青烟草花叶病毒 MP 由 268 个氨基酸组成, 相对分子质量 30 109.93, 等电点 8.77。三叶青烟草花叶病毒 RNA 聚合酶由 474 个氨基酸组成, 相对分子质量 54 639.11, 等电点 7.52。三叶青烟草花叶病毒带电蛋白由 40 个氨基酸组成, 相对分子质量 4 845.84, 等电点 11.89。

### 3.9 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白亲疏水性分析

三叶青烟草花叶病毒 CP、MP、RNA 聚合酶

和带电蛋白亲疏水性分析见图 6。高峰值(正值)的区域表示疏水的区域, 而负值的“低谷”区域是亲水区域。三叶青烟草花叶病毒 CP、MP、RNA 聚合酶为亲水性蛋白质, 而带电蛋白为疏水性蛋白质。

### 3.10 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白二级结构分析

三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白的二级结构见图 7。三叶青烟草花叶病毒 CP 二级结构由  $\alpha$  螺旋 (35.22%)、 $\beta$ -片层 (16.35%)、无规则卷曲 (48.43%) 构成; 三叶青烟草花叶病毒 MP 的二级结构预测如下: GOR 预测显示其二级结构由  $\alpha$  螺旋 (30.97%)、 $\beta$ -片层 (24.63%)、无规则卷曲 (44.40%) 构成; 三叶青烟草花叶病毒 RNA 聚合酶的二级结构预测如下: GOR 预测显示其二级结构由  $\alpha$  螺旋 (32.70%)、 $\beta$ -片层 (20.68%)、无规则卷曲 (46.62%) 构成; 三叶青烟草花叶病毒带电蛋白的二级结构预测如下: GOR 预测显示其二级结构由  $\beta$ -片层 (47.50%) 和规则卷曲 (52.50%) 构成。

CP 基因 ATGCTTACAATATCACTACTCCATCTCAGTTCGTGTTCTTGTTCATCAGCGTGGGC  
 AGCCCAATAGAGTTAATTAATTTATGTACTAATGCCTTAGGAAATCAGTTTCAA  
 ACACAACAAGCTAGAACTGTCGTTCAAAGACAATTGAGTGGTGTGGAAACCT  
 TCACCACAAGTAACTGTTAGGTTCCCTGATAGTGACTTTAAGGTGTACAGGTACA  
 ATGCGGTATTAGACCCGCTAGTCACAGCACTGTTAGGTGCATTGACACTAGAAA  
 TAGAATAATAGAAGTTGAAAATCAGGCGAACCCACGACTGCCGAAACGTTAGA  
 CGCTACTCGTAGAGTAGACGACGCAACGGTGGCCATAAGGAGCGCTATTAATA  
 TTTAGTAGTAGAATTGATCAGAGGAACCGATCTTATAATCGGAGCTCTTCGAG  
 AGCTCTCAGGTTTGGTTGGACCTCTGGTCTGCAACTGA。

MP ATGGCTCTAGTTGTTAAAAGGAAAAGTGAATATCAATGAGTATTATTGACCTGACAA  
 AATGGAGAAGATCTACCGTCGATGTTACCCTGTAAAGAGTGTATGTGTTT  
 CAAAGTTGATAAAAATAAGTTTCATGAGAATGAGTCAATGTCAGAGGTAACCTT  
 CTAAAGGAGTTAAGCTTATTGATAGTGGATACGCTGTTTAGCCGTTTGGTGT  
 CACGGGCGAGTGGAACTTGCCTGATAATTGCAGAGGAGGTGTGAGCGTGTGCT  
 GGTGGACAAAAGGATGGAAAGAGCCGACGAGGCCACTCTCGGATCTTATTACAC  
 AGCAGCTGCAAGAAAAGATTTGAGTCAAGGTCGTTCCCAATTATGCTATAACC  
 ACCCAGGACGCGATGAAAAACGCTGGCAAGTTTTAGTCAATATTAGAAATGTA  
 AAGATGTCAGCGGTTTCTGTCCGTTTCTCTGGAGTTTGTGTGCGGTGTATCGT  
 TTATAGAAAATAATAAAAATTAGGTTTGGAGAGAAAGATTACAAACGTGAGAGA  
 TGGAGGGCCCATGGAACCTACAGAAGAAGTTGTTGATGAGTTCATGGAAGATGT  
 CCCTATGTCAATCAGGCTTGCAAAGTTTCGATCTCGAACCGAAAAAAGAGTGAT  
 GTCCGTAAGGAAAATTAGTAGTAGTGATCGGTCAATGCCGAACAAGAACTAT  
 AGAAAATGTTAAGGATTTGGAGAAAATGAGTTTTAAAAAGAATAATTAATCGATG  
 ATGATTCGGAGACTACTGTCGCCGAATCGGATTCGTTTTAA。

RNA 聚合酶 ATGCAGTTTTACTATGATAAGTGTCTCCAGGCAACAGCACCATGATGAATAATT  
 TTGATGCTGTACCATGAGGTTGACAGACATTCATTGAATGTCAAAGATTGCAT  
 ATTGGATATGTCTAAGTCTGTTCTGCGCCTAAGGATCAAATCAAACCACTAATA  
 CCTATGGTACGAACGCCGGCAGAAATGCCACGCCAGACTGGACTATTGAAAAAC  
 TTAGTGGCGATGTTTAAAAGGAACGTTAACGCACCCGAGTTGTCTGGCATCATTG  
 ATATTGAAAATACTGCATCTTTAGTTGTAGATAAGTTTTTTGATAGTTATTGCTT  
 AAAGAAAAAGAAAACCAATAAAAAATGTTCTTTGTTCAAGTAGAGAGTCTCTC  
 AATAGATGGTTGAAAAGCAAGAACAGGTAACAATAGGCCAGCTCGCAGATTTT  
 GATTTGTGGATCTGCCAGCAGTTGATCAATACAGACACATGATCAAAGCACAAC  
 CCAAGCAAAAAGTTGGCACTTCGATCCAAACGGAGTACCCGGCTTTGCAGACGA  
 TTGTGTACCATTCAAAAAAGATCAATGCAATATTTGGCCCGTTGTTTGTAGTGGCT  
 ACTAGGCAATTACTGGACAGTGTGATTCGAGCAGATTTTTGTTTTTCAAGAA  
 AGACACCAGCGCAGATTGAGGATTTCTTCGGAGATCTCGACAGTCATGTGCCGAT  
 GGATGTCTTGGAGCTGGATATATCAAAAATACGACAAATATCAGAATGAATCCAC  
 TGTGCAGTAGAATACGAGATCTGGCGAAGATTGGGTTTTGAAGACTTCTTGGGAG  
 AAGTTTGGAAACAAGGGCATAGAAAAGACCCTCAAGGATTATACCGCAGGTA  
 TAAAAACTCGCATCTGGTATCAAAGAAAGAGCGGGGACGTCACGACGTGCATTG  
 GAAACACTGTGATCATTGCTCCATGTTTGGCCCGATGCTTCCGATGGAGAAAAT  
 AATCAAAGGAGCCTTTTGGCGTGACGATAGTCTGCTGAACCTTCCAAAGGGTGT  
 GTGTTCCGGATGTGCAACACTCCGCGAATCTATGTGGAATTTTGAAGCAAAAC  
 TGTTTTAAAAACAGTTTGGATACTTTTGGCGAAGATATGTAATACATCAGCAG  
 AGGATGCATTGTGATTACGATCCCTAAAGCTTATCTCGAAACTTGGTGCTAAA  
 CACCTCAAGTATTGGGAACACTTGGAGGAGTTCAGAATGTCTCTTTGTGATGTTG  
 CTGTTTCGTTGAACAATTGTGCGTATTACACACAGTTGGACGACGCTGTATGGGA  
 GGTTATAAGACCGCCCTCCAGGTTCTTTGTTTAAAAAGTCTGGCGAAGTAT  
 TTGCTGATAAAGTTCTTTTTAGAAGTTGTTTATAGATGGCTCTAGTTGTTAA。

带电蛋白 ATGATGATTCGGAGACTACTGTCGCCGAATCGGATTCGTTTTAAATATGCTTAC  
 AATACTACTCCATCTCAGTTCGTGTTCTTGTTCATCAGCGTGGGCAGACCCAAT  
 AGAGTTAATTA。

图 4 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白基因碱基组成

Fig. 4 Base composition of CP, MP, RNA polymerase and charged protein genes of tobacco mosaic virus in *T. hemsleyanum*

A  
 MSYNITTPSQFVFLSSAWADPIELINLCTNALGNQFQTQQARTVVQRQFSEVWKPSQVTV  
 RFPDSDFKVYRYNAVLDPVLTALLGAFDTRNRRIIEVENQANPTTAETLDAARRVDDATVAIR  
 SAINNLVVELIRGTGSYNRSSFESSGLVWVTSGPAT<sub>4</sub>

B  
 MALVVKGKVNINEFIDLTKMEKILPSMFTPVKSVCMCSKVDKIMVHENESLSEVNLLKGVK  
 LIDSGYVCLAGLVVTGEWNLDPNCRGGVSVCLVDKRMERADEATLGSSYYTAAAKKRFQF  
 KVVPNYAITQDAMKNVWQVLVNIIRNVKMSAGFCPLSLEFVSVCIVYRNLIKGLREKIT  
 NVRDGGPMELTEEVVDEFMEDVPMISIRLAKFRSRTGKKSVDVRKGISSSDRSMPNKNYR  
 NVKDFGEMSFKKNLIDDDSETTVAESDSF<sub>4</sub>

C  
 MQFYDYDKCLPGNSTMMNFDVAVTMRLLDISLNVKDCILDMSKSVPAKDKQIKPLIPMVRT  
 PAEMPRQTGLENLVAMFKRNVNAPELSGIIDIENTASLVVDKFFDSYLLKEKRKPNKNVS  
 LFSRESLNRWLEKQEQTIGQLADDFVLDPAVDQYRHMIIKAQPKQKLDTSIQTEYPALQ  
 TIVYHSHKINAIFGPLFSELTRQLLDSVDSRFLFFTRKTPAQIEDFFGLDLSHVPMDVLELD  
 ISKYDKYQNEFHCAVEYEIWRRLGFEDFLGEVWKQGHRKTTLKDYTAGIKTRIWIYQRKS  
 GDVTTICIGNTVIIAPCLAAMLPEKIIKGAFCGDDSLNFPKGCVPDPVQHSANLMWNFE  
 AKLFFKKQFGYFCGRYVIHHRGCVIYYDPLKLISKLGAKHLKYWEHLEEFRMSLCDVAVS  
 LNNCAYYTQLDDAVWEVHKTAPPGSFVYKSLAKYLSDKVLFRLSLFIDGSSC<sub>4</sub>

D  
 MMIRRLSPNRIRFKYVLQYHYSISVRVLVISVGRPNRVN<sub>4</sub>

图 5 三叶青烟草花叶病毒 CP (A)、MP (B)、RNA 聚合酶 (C) 和带电蛋白 (D) 氨基酸序列  
 Fig. 5 Amino acid sequence of CP (A), MP (B), RNA polymerase (C) and charged protein (D) of tobacco mosaic virus in *T. hemsleyanum*

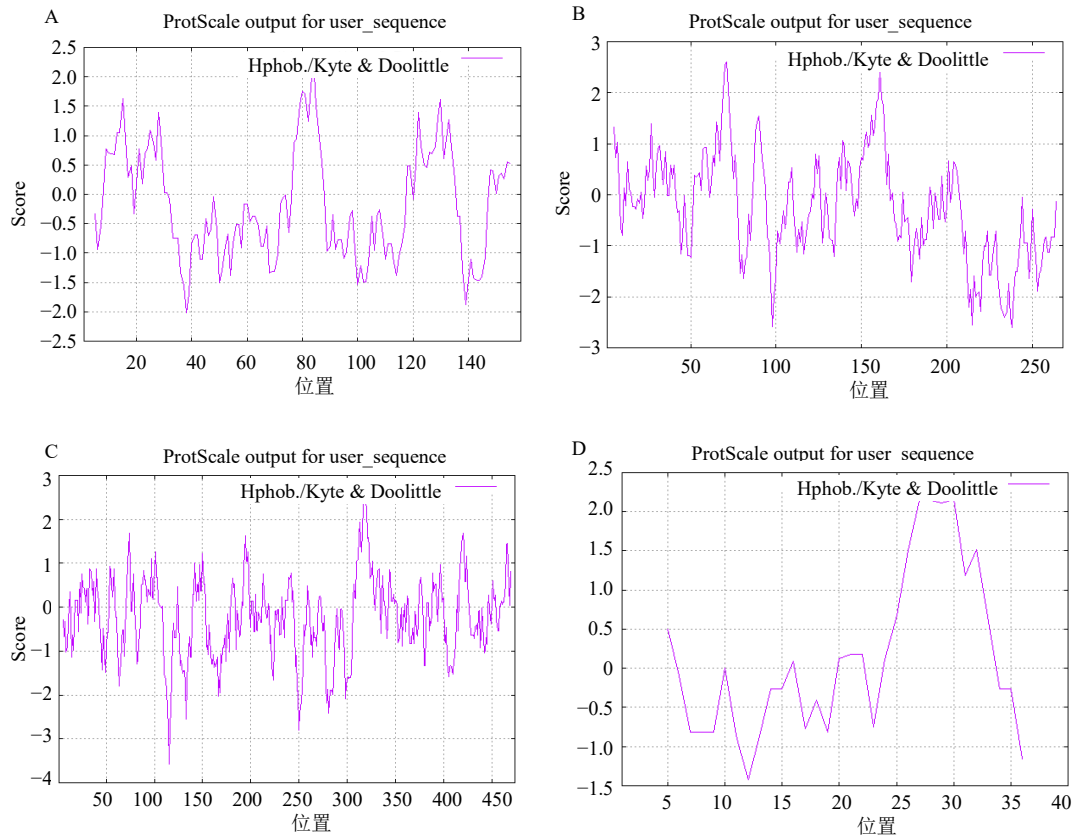
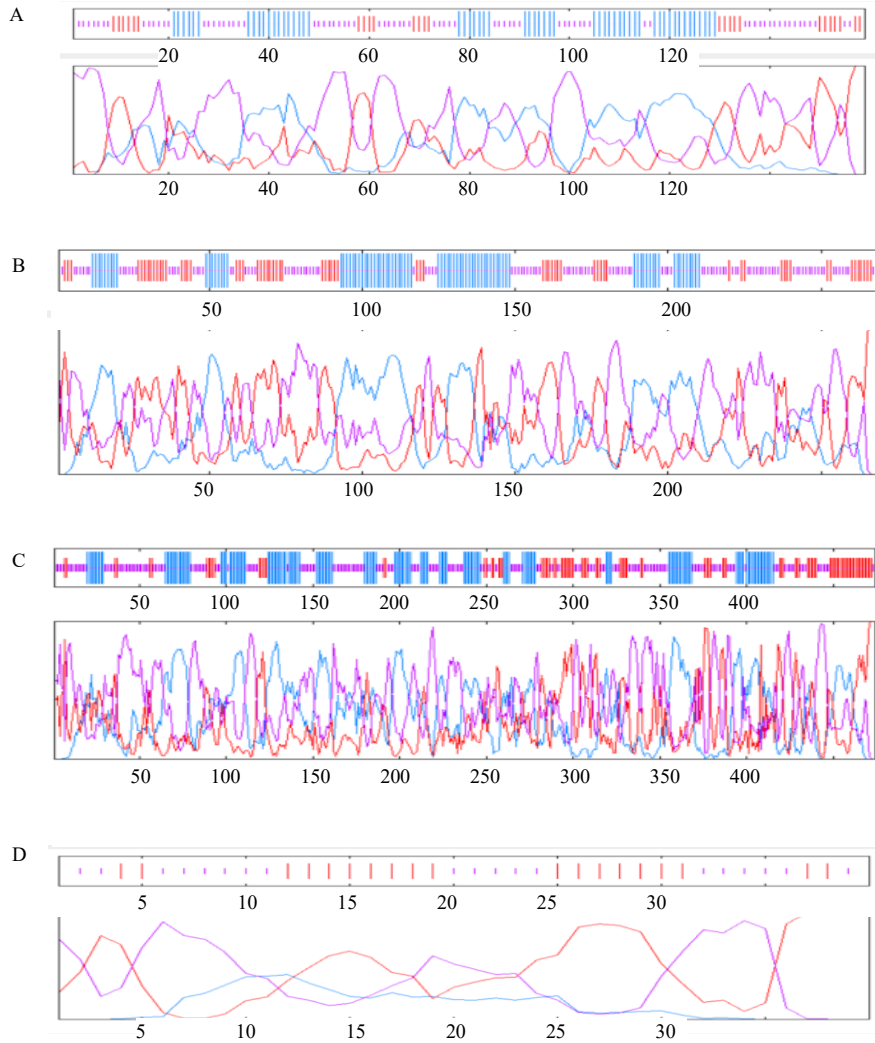


图 6 三叶青烟草花叶病毒 CP (A)、MP (B)、RNA 聚合酶 (C) 和带电蛋白 (D) 亲疏水值分布  
 Fig. 6 Distribution of hydrophilic and hydrophobic water values of CP (A), MP (B), RNA polymerase (C) and charged protein (D) of tobacco mosaic virus in *T. hemsleyanum*



蓝色代表  $\alpha$ -螺旋；红色代表  $\beta$ -片层；紫色代表不规则卷曲。

blue represents alphahelix; red represents extendedstrand; purple represents randomcoil.

图7 三叶青烟草花叶病毒 CP(A)、MP(B)、RNA 聚合酶 (C) 和带电蛋白 (D) 二级结构各部分的比例

Fig. 7 Proportion of secondary structure of CP (A), MP (B), RNA polymerase (C) and charged protein (D) of tobacco mosaic virus in *T. hemsleyanum*

### 3.11 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白三级结构分析

三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白三级结构 SWISS-MODEL 预测分析见图 8。三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白三级结构均为单体。

### 3.12 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白亚细胞定位

三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白基因表达部位 WoLFPSort 预测分析见图 9。三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白分别定位于叶绿体、叶绿体、细胞质和线粒体中。

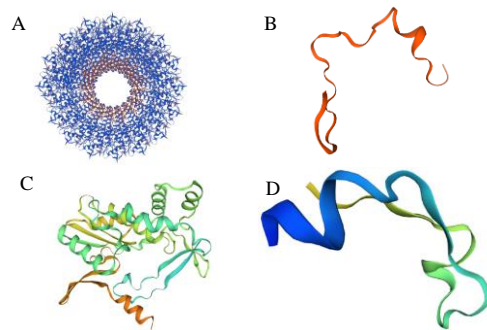
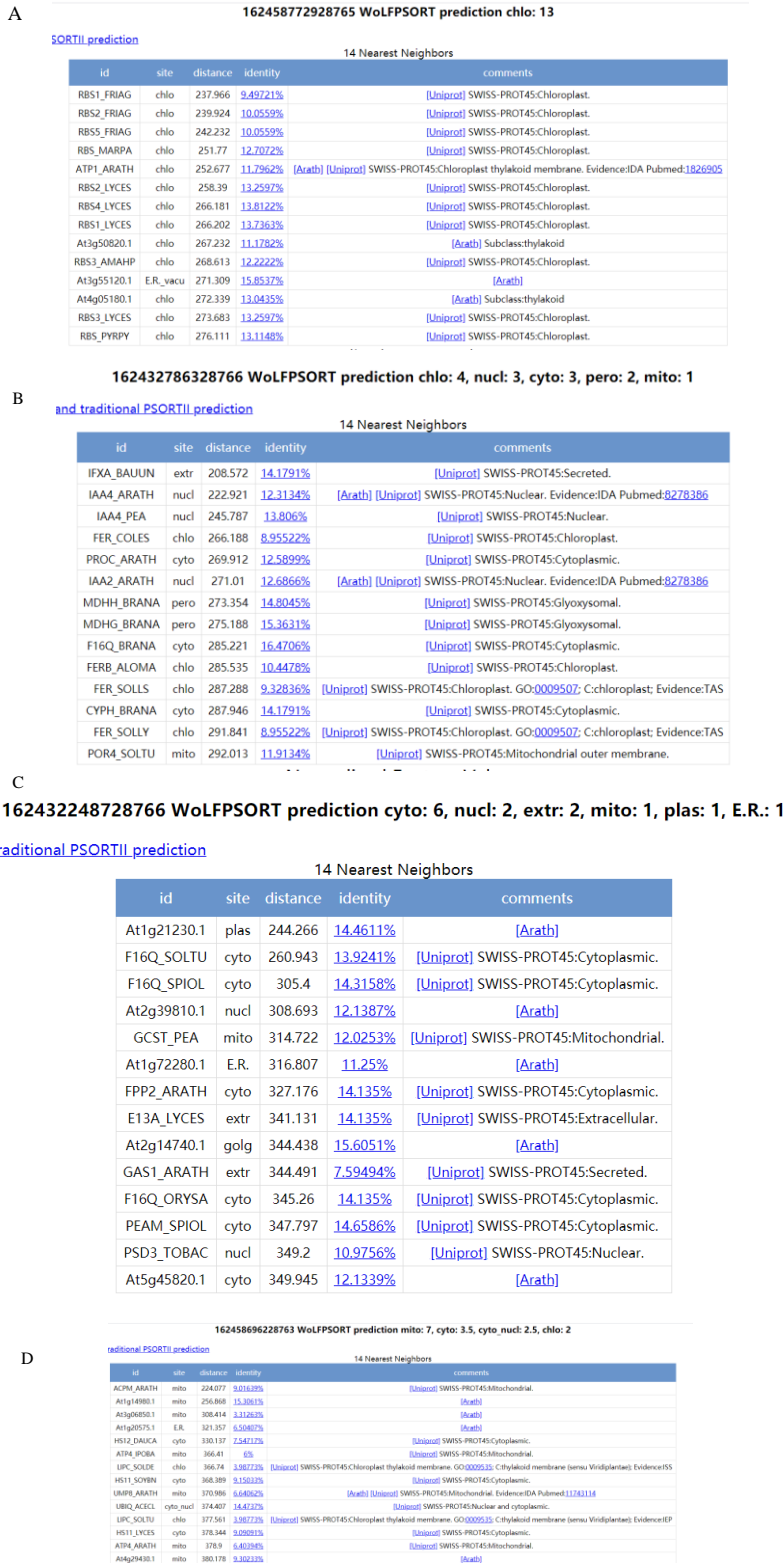


图8 三叶青烟草花叶病毒 CP(A)、MP(B)、RNA 聚合酶 (C) 和带电蛋白 (D) 三级结构

Fig. 8 Tertiary structure of CP (A), MP (B), RNA polymerase (C) and charged protein (D) of tobacco mosaic virus in *T. hemsleyanum*





Chlo-叶绿体; Plas-质膜; Vacu-液泡膜; Nucl-细胞核; Cyto-细胞质; E.R-内质网; extr-细胞外; mito-线粒体。

Chlo-chloroplast; Plas-plasma membrane; Vacu-vacuole membrane; Nucl-nucleus; Cyto-cytoplasm; E.R-endoplasmic reticulum; extr-extracellular;

Mito-mitochondria.

图 9 三叶青烟草花叶病毒 CP (A)、MP (B)、RNA 聚合酶 (C) 和带电蛋白 (D) 亚细胞定位  
 Fig. 9 Subcellular localization of CP (A), MP (B), RNA polymerase (C) and charged protein (D) of tobacco mosaic virus in *hemsleyanum*

### 3.13 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白系统进化分析

从构建的进化树（图 10）中可见，三叶青烟草花叶病毒 CP 与烟草花叶病毒 CP 在一个分支下，这说明三叶青烟草花叶病毒 CP 在进化上与烟草花叶病毒 CP 的亲缘关系较近，同源率为 99.58%；三叶青烟草花叶病毒 MP 与烟草花叶病毒 MP 在一个分支下，这说明三叶青烟草花叶病毒 MP 在进化上与烟草花叶病毒 MP 的亲缘关系

较近，同源率为 99.63%；三叶青烟草花叶病毒 RNA 聚合酶与烟草花叶病毒 RNA 聚合酶在一个分支下，这说明三叶青烟草花叶病毒 RNA 聚合酶在进化上与烟草花叶病毒 RNA 聚合酶的亲缘关系较近，同源率为 98.04%；三叶青烟草花叶病毒带电蛋白与烟草花叶病毒带电蛋白在一个分支下，这说明三叶青烟草花叶病毒带电蛋白在进化上与烟草花叶病毒带电蛋白的亲缘关系较近，同源率为 100%。

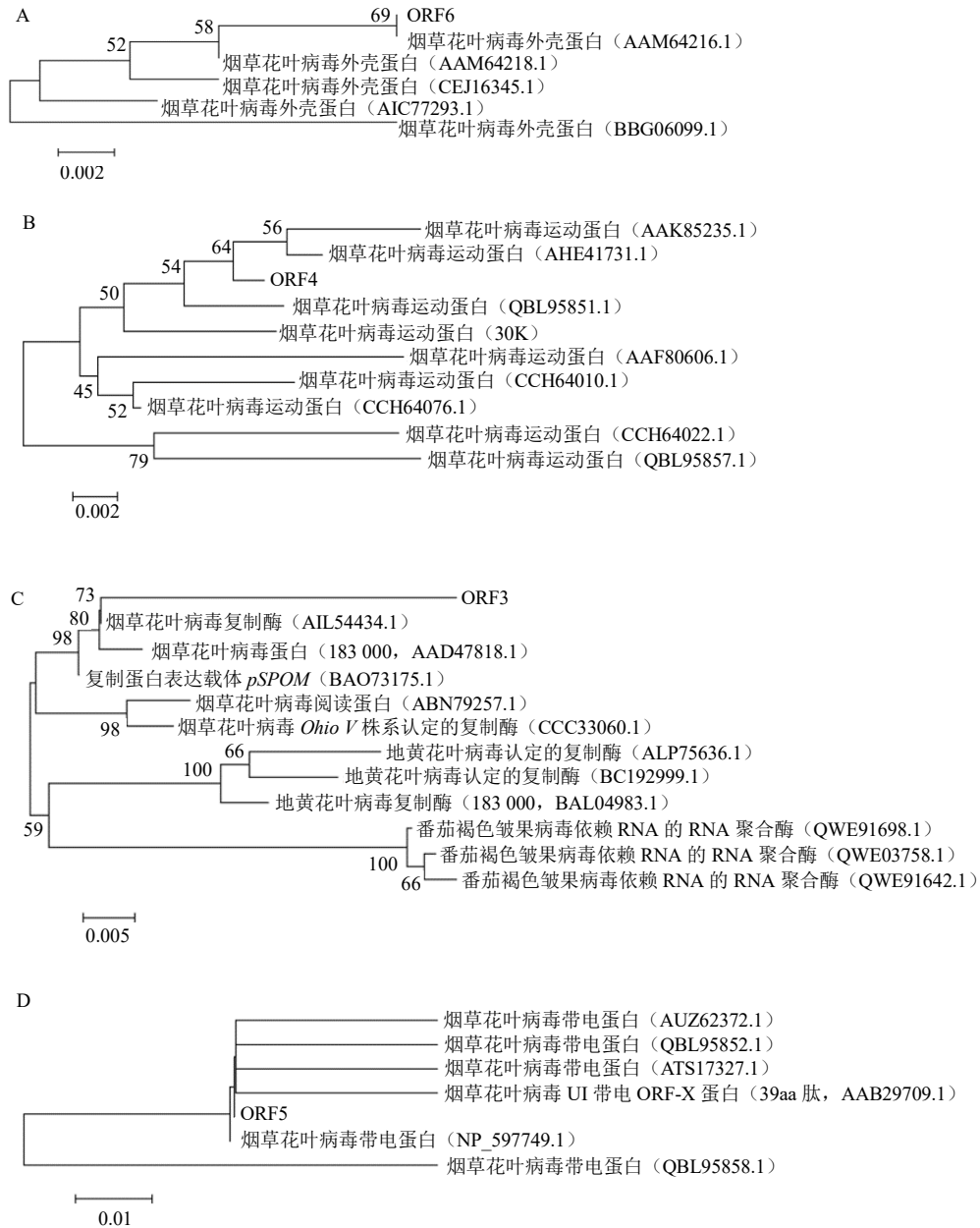


图 10 三叶青烟草花叶病毒 CP (A)、MP (B)、RNA 聚合酶 (C) 和带电蛋白 (D) 系统进化分析

Fig. 10 Phylogenetic analysis of CP (A), MP (B), RNA polymerase (C) and charged protein (D) of tobacco mosaic virus in *T. hemsleyanum*

### 3.14 三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白的序列比对信息

三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白同源蛋白的序列比对信息见图 11。图

11 中“\*”号区域是该蛋白家族的保守结构域。由图 11 可知，三叶青烟草花叶病毒 CP、MP、RNA 聚合酶和带电蛋白与烟草花叶病毒保守结构域较为丰富。

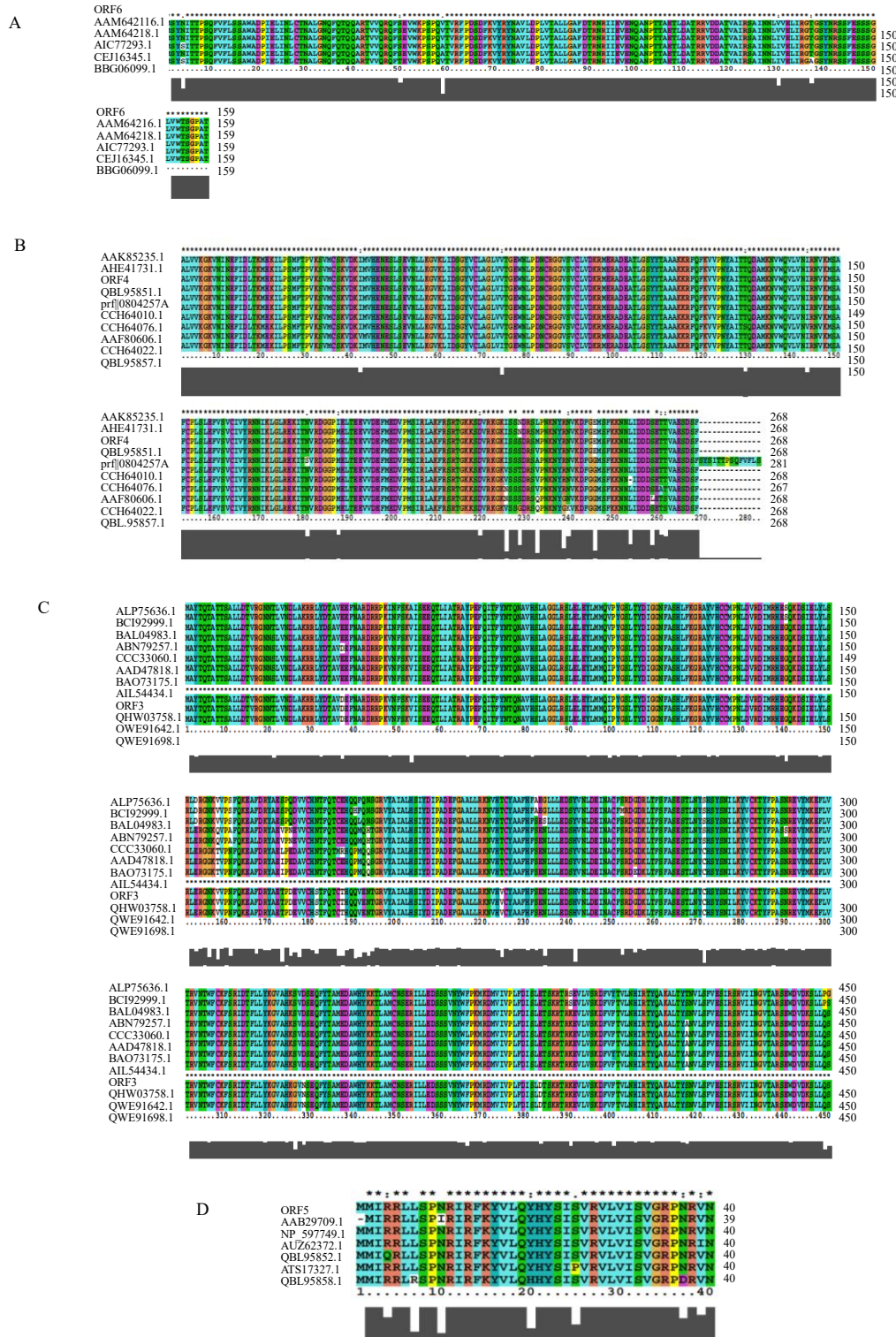


图 11 三叶青烟草花叶病毒 CP (A)、MP (B)、RNA 聚合酶 (C) 和带电蛋白 (D) 氨基酸序列的同源性比较  
Fig. 11 Homology comparison of amino acid sequences of CP (A), MP (B), RNA polymerase (C) and charged protein (D) of tobacco mosaic virus in *T. hemsleyanum*

#### 4 讨论

烟草花叶病毒是影响世界农业生产的重要植物病毒,给农业生产造成了巨大的经济损失,其检测和防治工作一直备受关注。史跃伟等<sup>[17]</sup>通过 RT-PCR 的方法获得了贵州烟草花叶病毒 CP 基因,该基因全长为 507 bp,并利用原核表达技术表达了贵州烟草花叶病毒 CP 蛋白,为后期单克隆抗体的制备奠定了基础。烟草花叶病毒南瓜分离物 CP 基因与 GenBank 上其他 TMV 分离物 CP 基因核苷酸序列的同源性为 86.5%~99.0%,推导的氨基酸序列同源性为 93.7%~99.4%,烟草花叶病毒南瓜分离物 CP 基因与原核表达载体 pET-22b (+) 连接,在大肠杆菌 BL21 (DE3) pLysS 诱导表达出相对分子质量约 20 000 的融合蛋白<sup>[18]</sup>。云南 5 个不同地区烟草花叶病毒 CP 基因均为 480 个碱基,编码 159 个氨基酸,与 TMV-U1、P、B 株和福建分离物有较高的同源率(90%以上),而与 TMV 中国株系的同源率较低(70%~80%),与 TMV 中国株应属于同种病毒不同的株系<sup>[19]</sup>。李凡等<sup>[20]</sup>的研究结果也证实了这一观点,再次确认烟草花叶病毒云南分离物 CP 基因为 480 个碱基,编码 159 个氨基酸,并指出烟草花叶病毒云南分离物 CP 基因与 TMV-U1 株系和 TMV 韩国普通株系核苷酸同源性均为 100%。但也有研究表明,云南烟草花叶病毒 CP 基因含 477 个核苷酸,编码 159 个氨基酸<sup>[21]</sup>。本实验结果也表明,三叶青烟草花叶病毒 CP 基因 cDNA 总长度为 480 bp, G+C 含量为 42.71%,编码 159 个氨基酸组成;三叶青烟草花叶病毒 CP 为亲水性蛋白质,二级结构由  $\alpha$  螺旋(35.22%)、 $\beta$ -片层(16.35%)、无规则卷曲(48.43%)构成,三级结构为单体,定位于叶绿体,与烟草花叶病毒 CP 亲缘关系较近,同源性为 99.58%。

烟草花叶病毒的致病性是由 CP 基因及 3'端非编码区以外的基因(如 MP 基因、复制酶基因等)决定的。烟草花叶病毒 MP 基因决定植物病毒在寄主植物细胞间的移动从而可决定病毒对寄主的侵染性。有研究表明,烟草花叶病毒蚕豆分离 MP 基因由 807 个核苷酸组成,编码 268 个氨基酸,烟草花叶病毒的 MP 基因在不同株系中非常保守,同源性极高<sup>[22]</sup>。王得元等<sup>[23]</sup>对广东烟草花叶病毒 MP 基因进行 PCR 扩增,获得了 1 条 0.8 kb 的特异扩增产物,与已发表的 TMV MP 基因编码区 803 bp 大小相近。余晓红等<sup>[24]</sup>克隆了烟草花叶病毒 MP 基因,

发现 MP 基因含有 1 个 507 bp 的阅读框架,编码 268 个氨基酸,与国外发表的 TMVU1 株运动蛋白基因 cDNA 的核苷酸序列及推导的氨基酸序列相比,同源率均高达 98 以上,再次表明烟草花叶病毒 MP 基因在同一病毒的不同株系中是相当保守的。本实验结果表明,三叶青烟草花叶病毒 MPcDNA 总长度为 807 bp, G+C 含量为 40.40%,编码 268 个氨基酸,为亲水性蛋白质;三叶青烟草花叶病毒 MP 二级结构由  $\alpha$  螺旋(30.97%)、 $\beta$ -片层(24.63%)、无规则卷曲(44.40%)构成,三级结构为单体;三叶青烟草花叶病毒 MP 定位于叶绿体,与烟草花叶病毒 MP 亲缘关系较近,同源性为 99.63%。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] 李士敏,刘京骅,杨颖欣,等. 三叶青 HPLC 指纹图谱研究及多成分含量测定 [J]. 亚太传统医药, 2022, 18(7): 46-51.
- [2] Ji T, Ji W W, Wang J, et al. *Tetrastigma hemsleyanum* alleviates sarcoidosis through metabolomic regulation and Th17/Treg immune homeostasis [J]. *J Funct Foods*, 2022, 88: 104910.
- [3] 洪森荣,刘雯莉,宋冰雁,等. 怀玉山产三叶青叶绿体基因组特征及其系统进化关系 [J]. 中草药, 2023, 54(16): 5358-5371.
- [4] 尹明华,曾淑琴,曾炎,等. 怀玉山三叶青烟草病毒增殖蛋白 2A 基因克隆和功能分析 [J]. 种子, 2022, 41(3): 45-51.
- [5] 林国卫,闻静,石光禹,等. 侵染怀玉山三叶青的病毒 RT-PCR 鉴定 [J]. 分子植物育种, 2020, 18(3): 968-975.
- [6] 尹明华,胡嘉琦,胡明艳,等. 怀玉山三叶青 TOM2B isoform X1 基因克隆和功能分析 [J]. 山东农业科学, 2021, 53(8): 1-10.
- [7] 尹明华,陈佳雯,陈瑞,等. 怀玉山三叶青烟草病毒增殖蛋白 3 基因克隆和功能分析 [J]. 江苏农业科学, 2022, 50(6): 32-41.
- [8] Gao F L, Liu X W, Du Z G, et al. Bayesian phylodynamic analysis reveals the dispersal patterns of tobacco mosaic virus in China [J]. *Virology*, 2019, 528: 110-117.
- [9] 黄倩茹,牛永浩,吴宽,等. 陕西省草莓病毒种类的 LncRNA 测序初步鉴定 [J]. 园艺学报, 2021, 48(8): 1589-1594.
- [10] 訾豪然,许海顺,吴盛楠,等. 不同产地林下三叶青多糖指纹图谱研究 [J]. 中国农学通报, 2021, 37(27): 139-145.
- [11] 丁丽,王淑君,徐冰婉,等. 基于体外抗氧化试验的三叶青活性物质提取工艺的研究 [J]. 食品科技, 2021,

- 46(12): 224-229.
- [12] 胡竹元, 余志怡, 徐斌. 三叶青总黄酮增强三阴性乳腺癌细胞对 EGFR-TKI 敏感性的机制研究 [J]. 中国临床药理学杂志, 2022, 38(8): 772-776.
- [13] 刘荣昌, 徐红兵, 余家中, 等. 3 种不同容器三叶青育苗技术试验初报 [J]. 浙江农业科学, 2022, 63(4): 705-707.
- [14] 杜家会, 李振斌, 姚天月, 等. 贵州黔西南州野生三叶青种质资源调查分析 [J]. 安徽农学通报, 2021, 27(17): 45-46.
- [15] 屠玲艳, 吴学谦, 许海顺. 水分胁迫对三叶青叶绿体超微结构及黄酮合成关键酶的影响 [J]. 浙江农林大学学报, 2021, 38(3): 577-586.
- [16] 陆飞彪, 陈妍月, 程汝滨, 等. 三叶青愈伤组织诱导增殖体系建立及天然有机物促进总黄酮合成的研究 [J]. 时珍国医国药, 2019, 30(12): 2987-2990.
- [17] 史跃伟, 郭玉双. 烟草花叶病毒贵州分离物 CP 基因的原核表达及纯化 [J]. 贵州农业科学, 2012, 40(12): 18-20.
- [18] 刘金亮, 王凤婷, 魏毅, 等. 烟草花叶病毒贵州分离物 CP 基因的原核表达及纯化 [J]. 华北农学报, 2010, 25(5): 6-10.
- [19] 丁铭, 方琦, 张丽珍, 等. 云南五个不同地区烟草花叶病毒外壳蛋白基因的序列比较 [J]. 中国病毒学, 2004, 19(1): 76-78.
- [20] 李凡, 周雪平, 陈海如, 等. 云南五个不同地区烟草花叶病毒外壳蛋白基因的序列比较 [J]. 浙江大学学报(农业与生命科学版), 2000, 26(3): 261-265.
- [21] 王芳, 鄢波, 王玲仙, 等. 云南烟草花叶病毒外壳蛋白基因的分离克隆及序列分析 [J]. 西南农业学报, 1999, 12(2): 7-11.
- [22] 陈青. 烟草花叶病毒蚕豆分离物移动蛋白基因克隆及序列分析 [J]. 浙江农业大学学报, 1999, 25(3): 239.
- [23] 王得元, 刘志昕, 潘俊松, 等. 广东烟草花叶病毒移动蛋白基因的 cDNA 克隆和酶切鉴定 [J]. 江西农业大学学报, 1998, 20(3): 334-242.
- [24] 余晓红, 朱玉贤, 应华澄, 等. 烟草花叶病毒运动蛋白基因的 cDNA 克隆、序列测定及植物转化 [J]. 生物工程学报, 1996, 12(2): 211-214.

[责任编辑 时圣明]