

中图分类号: S532 文献标识码: A 文章编号: 1672-3635(2024)01-0001-09

DOI: 10.19918/j.cnki.1672-3635.2024.01.001

遗传育种

马铃薯转录组数据中的病毒序列挖掘

陈家茹¹, 杨曼华¹, 陈汝豪¹, 孟繁烨¹, 余涛¹, 柴婷婷², 许彬燕¹, 聂碧华^{1*}

(1. 果蔬园艺作物种质创新与利用全国重点实验室/农业农村部马铃薯生物学与生物技术重点实验室,

华中农业大学, 湖北 武汉 430070; 2. 湖北省农业技术推广总站, 湖北 武汉 430070)

摘要: 转录组测序数据中包含寄主所感染的病毒序列信息。为分析和验证马铃薯转录组数据中存在的病毒序列, 试验收集了关于马铃薯块茎发育、抗病、抗逆等研究的9个马铃薯转录组测序数据样本, 过滤掉来源于马铃薯转录本的读段, 组装并与病毒数据库进行本地比对, 获得样本中相关病毒序列的种类和数量信息。病毒来源的读段在各样本中普遍存在, 占总读段的0.12%~20.24%。这些读段组装得到10种植物病毒, 包括9种RNA病毒和1种DNA病毒。其中来源于马铃薯S病毒(Potato virus S, PVS)的序列最多, 组装得到28条近全长序列。进一步系统进化分析表明, 这些样品中同时存在PVS^A和PVS^O两种株系类型。此外, 通过RT-PCR对转录组数据原始样本中的马铃薯H病毒(Potato virus H, PVH)、马铃薯卷叶病毒(Potato leafroll virus, PLRV)、马铃薯M病毒(Potato virus M, PVM)、PVS、马铃薯X病毒(Potato virus X, PVX)和马铃薯Y病毒(Potato virus Y, PVY)6种病毒进行检测, 结果显示RT-PCR验证结果与测序数据组装结果基本一致。该研究为马铃薯病毒多样性和进化分析提供一种高通量测序的方法。

关键词: 马铃薯; 病毒; 高通量测序; 转录组; RT-PCR

Identification of Virus Sequences in Potato Transcriptome Data

CHEN Jiaru¹, YANG Manhua¹, CHEN Ruhao¹, MENG Fanye¹, YU Tao¹, CHAI Tingting², XU Binyan¹, NIE Bihua^{1*}

(1. National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops/Key Laboratory of Potato Biology and

Biotechnology, Ministry of Agriculture and Rural Affairs, Huazhong Agricultural University, Wuhan, Hubei 430070, China;

2. Agricultural Technology Extension Station of Hubei Province, Wuhan, Hubei 430070, China)

Abstract: Transcriptome sequencing data often contain valuable information about viruses infecting the host. In order to analyze and validate the presence of virus sequences in potato transcriptome data, nine transcriptome sequencing data samples related to potato tuber development, disease resistance, and stress tolerance were collected. After filtering out reads originating from potato transcripts, the remaining unmapped reads were assembled and locally aligned against a virus database to obtain information on the types and quantities of virus sequences in these samples. The results revealed that reads originating from viruses were commonly present in all samples, accounting for 0.12% to 20.24% of the total reads. These reads were assembled into sequences representing 10 different plant viruses,

收稿日期: 2024-02-19

基金项目: 国家自然科学基金(31971989); 国家马铃薯产业技术体系(CARS-09-P07)。

作者简介: 陈家茹(1999-), 女, 硕士研究生, 研究方向为马铃薯遗传育种。

*通信作者(Corresponding author): 聂碧华, 博士, 副教授, 研究方向为马铃薯遗传育种, E-mail: nbihua@mail.hzau.edu.cn。

including nine RNA viruses and one DNA virus. Among them, sequences derived from potato virus S (PVS) were the most abundant, with 28 nearly full-length sequences assembled. Further phylogenetic analysis indicated the coexistence of two strain types, PVS^A and PVS^O, in these samples. Additionally, reverse transcription-polymerase chain reaction (RT-PCR) was performed to detect six viruses including potato virus H (PVH), potato leafroll virus (PLRV), potato virus M (PVM), potato virus S (PVS), potato virus X (PVX), and potato virus Y (PVY) in the original transcriptome data samples. The RT-PCR validation results were generally consistent with the assembly results from the sequencing data. Overall, this study provides a high-throughput approach for analyzing the diversity and evolution of viruses in potatoes.

Key Words: potato; virus; high-throughput sequencing; transcriptome; RT-PCR

马铃薯是中国重要的粮食作物之一, 对保障粮食安全具有重要意义^[1]。病毒病是马铃薯生产上重要病害, 常见马铃薯病毒主要有马铃薯Y病毒(Potato virus Y, PVY)、马铃薯卷叶病毒(Potato leafroll virus, PLRV)、马铃薯X病毒(Potato virus X, PVX)、马铃薯A病毒(Potato virus A, PVA)、马铃薯S病毒(Potato virus S, PVS)、马铃薯M病毒(Potato virus M, PVM)、烟草脆裂病毒(Tobacco rattle virus, TRV)和马铃薯帚顶病毒(Potato mop-top virus, PMTV), 以及马铃薯纺锤块茎类病毒(Potato spindle tuber viroid, PSTVd)^[2,3]。马铃薯被病毒侵染后出现植株矮化、花叶、卷曲等症状, 影响马铃薯产量, 且病毒逐年累积造成种薯退化, 严重影响马铃薯品质, 制约马铃薯产业发展^[4]。建立快速高效准确的病毒检测方法, 是马铃薯病毒防控首要任务。

人类基因组计划的完成标志着生命科学研究进入后基因组时代。2011年完成的马铃薯基因组序列测序为相关研究带来新机遇^[5-7]。随着测序技术不断更新应用, 产生大量转录组测序数据。转录组测序(RNA sequencing, RNA-seq)技术在获得动植物体转录本的同时, 也能获得样品中所感染病毒的序列信息, 因此转录组数据在完成寄主目标基因分析后, 还可进行病毒资源挖掘^[8]。在动物中, Sparks等^[9]对茶翅蜡(*Halyomorpha halys*)的转录组测序数据使用Trinity拼接组装, 发现一种新病毒, 命名为茶翅蜡病毒贝尔茨维尔分离物(*Halyomorpha halys* virus isolate Beltsville); Smith等^[10]从艺神袖蝶(*Heliconius erato*)的转录组数据中

发现和鉴定一种传染性软腐病病毒属(*Iflavirus*)的新成员, 命名为艺神袖蝶传染性软腐病病毒(*Heliconius erato* Iflavirus)。在植物中, Sun等^[11]通过转录组测序技术鉴定侵染梨的病原物, 首次在中国发现新疆库尔香梨(*Pyrus*)是啤酒花矮化类病毒(Hop stunt viroid, HSVd)的寄主; 李伟佳等^[12]对感病草莓进行转录组测序, 经比对发现组装结果中有12条序列可被注释为4种草莓病毒: 草莓坏死休克病毒(Strawberry necrotic shock virus, SNSV)、草莓轻型黄边病毒(Strawberry mild yellow edge virus, SMYEV)、草莓皱缩病毒(Strawberry crinkle virus, SCV)和草莓镶脉病毒(Strawberry vein banding virus, SVBV); 胡国君等^[13]基于苹果树small RNA(sRNA)深度测序的数据集, 鉴定到以苹果褪绿叶斑病毒(Apple chlorotic leaf spot virus, ACLSV)、苹果绿皱纹相关病毒(Apple green crinkle associated virus, AGCaV)、苹果茎痘病毒(Apple stem pitting virus, ASPV)为主的多种苹果病毒。这些研究表明RNA-seq数据可应用于病毒的发掘和鉴定。本研究收集了实验室积累的部分马铃薯RNA-seq数据, 通过对转录组数据中的病毒序列进行组装和与病毒数据库比对, 获得了样品中病毒相关信息, 结果对马铃薯病毒的检测和新病毒的发现有重要价值。

1 材料与方法

1.1 数据收集

本研究收集和使用的RNA-seq数据, 包括马铃薯低温糖化(Cold sweetening, CS)、块茎发育

(Tuber development/Illumination, TD; Tuberization/E26, TU)、休眠期(Domancy, DM)、抗晚疫病(Late blight T, LBT; Late blight W, LBW)、抗病毒(Potato virus, PV)、抗寒(Cold resistance, CR)、色素(Pigment, PG)等处理的测序数据, 合计9组, 总计约1 T数据。数据全部来自华中农业大学马铃薯实验室前期开展的相关马铃薯块茎发育、抗病、抗逆等研究。

1.2 生物信息数据分析方法

1.2.1 数据的质量控制

进行数据分析前, 对RNA-seq数据进行预处理: 使用FastQC(v0.11.3)软件对数据质量进行评估, 从结果中可得到各样本数据的Reads质量; 对于质量不合格的数据, 用Trimmomatic进行数据过滤, 设置过滤参数为SLIDINGWINDOW:5:20 LEADING:5 TRAILING:5 MINLEN:50, 得到Clean_data, 用于后续分析。

1.2.2 RNA-seq数据匹配马铃薯参考基因组

用TopHat(v2.0.10)软件参照马铃薯基因组PGSC_DM_v4.03_pseudomolecules(http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml)对Clean_data数据进行比对, 参数设置: 最小内含子长度为10, 最大内含子长度为15 000, 允许2个Mismatch^[4], 从而过滤掉来源于马铃薯转录本的Reads, 并统计无法比对上马铃薯基因组Unmapped reads的数量, 及占总Reads的比例。

1.2.3 使用Trinity软件进行组装

对包含Unmapped reads数据的文件(unmapped.bam), 使用Samtools(v0.1.18)软件根据BAM文件的Flag标签, 提取双端Reads, 得到unmapped_reads1.fastq和unmapped_reads2.fastq。使用命令如下:

```
samtools view unmapped.bam | awk '{if($2~/69/)print ">" $1"\n" $10"\n + \n" $NF}' > unmapped_reads1.fastq
```

```
samtools view unmapped.bam | awk '{if($2~/133/)print ">" $1"\n" $10"\n + \n" $NF}' > unmapped_reads2.fastq
```

获取双端Reads后, 使用Trinity(v2.1.0)软件进行序列重新组装。使用命令如下:

```
Trinity--seqType fq--max_memory 10G--left unmapped_reads1.fastq--right unmapped_reads2.fastq--CPU 10
```

1.2.4 病毒数据库构建

本研究使用2个病毒数据库: 第一个数据库的构建基于从NCBI公共数据库(<https://www.ncbi.nlm.nih.gov/genome/viruses/>)下载的106 472条病毒核苷酸序列, 用于对测序数据Reads进行病毒序列注释, 分析病毒Reads在总Reads所占比例, 命名为数据库I; 第二个数据库是从NCBI公共数据库下载的病毒标准序列数据库(<ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/>), 由于该数据库并未包括所有已报道的马铃薯病毒序列, 因此在该数据库加入目前已报道的对马铃薯有侵染能力的45种病毒和1种类病毒(PSTVd)全长序列, 用于挖掘侵染马铃薯的未知病毒, 以及分析组装得到的病毒序列多样性, 命名为数据库II。

1.2.5 本地化Blast与病毒数据库进行同源性比对

使用本地化Blast+(v2.2.31)工具将Unmapped reads数据与病毒数据库I比对, 进行病毒注释并统计病毒来源Reads数目及占全部Reads的比例; 将组装得到的Contig与病毒数据库II进行比对, 研究组装结果中病毒多样性, 并对E值 $< 1 \times 10^{-5}$ 进行统计。

1.2.6 系统进化树构建

选择组装结果中接近完整病毒基因组的Contig序列(即覆盖90%以上病毒基因组的序列), 与NCBI公共数据库里的病毒全长序列一起进行系统进化分析, 构建系统进化树; 将病毒序列导入MEGA(v6.06)软件, 使用软件自带的Clustal W工具进行多重序列比对。用MEGA(v6.06)软件通过邻接法(Neighbour-joining, NJ)构建系统进化树(参数: 自展检验Bootstrap method 1 000次)。

1.3 病毒检测引物和PCR反应体系

用于RT-PCR验证测序分析结果所用的病毒检测引物(表1)共12个。

本研究PCR反应体系(25 μ L)包括反应缓冲液2 $\times$ Taq Plus Master Mix 12.5 μ L, 9.5 μ L ddH₂O, 正向和反向特异性引物各1 μ L以及扩增模板的cDNA 1 μ L。反应程序为: 95 $^{\circ}$ C预变性3 min; 95 $^{\circ}$ C变性

30 s, 55 ℃ 退火 30 s, 72 ℃ 延伸 1 min, 共 37 个循环; 72 ℃ 延伸 10 min。

2 结果与分析

2.1 RNA-seq 数据中病毒序列的注释

为统计测序数据中的病毒序列信息, 将 RNA-seq 数据比对马铃薯参考基因组以过滤掉马铃薯基因组数据, 获得无法匹配马铃薯参考基因组的 Unmapped reads。结果显示(表 2), 各样本 RNA-seq 数据中 Unmapped reads 占总 Reads 的比

例为 18.90%~31.24%, 平均为 23.16%。进一步将 Unmapped reads 进行病毒注释, 结果表明, 病毒来源的 Reads(Virus reads)在各样本中普遍存在, 但占总 Reads 的比例差异较大。其中 CR、CS、LBT、LBW 4 个样品比例相对较低, 均小于 1%, LBT 样品病毒相关序列比例最低(0.12%), 仅占 Unmapped reads 数量的 0.48%; 而 DM、PG、PV、TD、TU 5 个样品占总 reads 比例较高, 为 2.42%~20.24%, 其中 DM 样品病毒相关序列比例最高(20.24%), 占 Unmapped reads 数量的 64.77%。

表 1 病毒检测引物

Table 1 Primers for virus detection

引物名称 Primer name	引物序列 Primer sequence	片段长度(nt) Fragment length	用途 Function
PLRV-4S	5'-CGCGCTAACAGAGTTCAGCC-3'	336	PLRV detection
PLRV-5A	5'-GCAATGGGGTCCAACTCAT-3'		
PVM-1S	5'-CGAAGGCTGTGTAGGCTGTATG-3'	226	PVM detection
PVM-2A	5'-CGATGTCTTTGTGCGTATTGTG-3'		
PVS-12S	5'-TGGCGAACACCGAGCAAATG-3'	425	PVS detection
PVS-12A	5'-ATGATCGAGTCCAAGGGCACTG-3'		
PVX-A11	5'-TCGAAAGATGTCAGGACCAGC-3'	226	PVX detection
PVX-AA11	5'-TGTGGGCACCTTCATGTCCTTC-3'		
PVY-3S	5'-ACGTCCAAAATGAGAATGCC-3'	480	PVY detection
PVY-4A	5'-TGCTGTTCTGTATGTGACCT-3'		
PVH-1S	5'-GCAATGAGGCTTAGTTCA-3'	310	PVH detection
PVH-2A	5'-CTTATGCGCATCTATCAATC-3'		

表 2 所有 RNA-seq 数据进行病毒注释后的数据

Table 2 Results of all RNA-Seq data BLAST virus nucleotide database

数据来源 Data	读长(nt) Reads length	读段总数 Number of total reads	未比对上读段数 Number of unmapped reads	未比对读段占总读段的比例(%) Ratio of unmapped reads to total reads	病毒读段数 Number of virus reads	病毒读段占总读段的比例(%) Ratio of virus reads to total reads
CR	150/125(Paired-end)	1 169 963 066	248 073 327	21.20	11 457 571	0.98
CS	150(Paired-end)	91 199 390	19 903 119	21.82	180 695	0.20
DM	125(Paired-end)	74 048 666	23 135 368	31.24	14 984 621	20.24
LBT	150(Paired-end)	574 151 104	139 231 186	24.25	664 326	0.12
LBW	150(Paired-end)	988 616 094	243 299 188	24.36	1 696 893	0.17
PG	150(Paired-end)	666 134 146	155 590 293	23.36	16 117 233	2.42
PV	150(Paired-end)	38 097 666	8 703 710	22.85	2 529 501	6.64
TD	150(Paired-end)	276 912 740	59 957 554	21.65	17 626 684	6.37
TU	90(Paired-end)	37 856 854	7 154 381	18.90	4 878 655	12.89
Total/Average		3 916 979 726	905 048 126	23.16	70 136 179	1.79

2.2 病毒序列的组装

本试验使用 Trinity 软件对 Unmapped reads 进行组装, 组装结果比对病毒数据库以研究样本中病毒种类。结果显示(表3), 在9个样本中组装到

10种植物病毒, 包括9种RNA病毒和1种DNA病毒, 共2 129条植物病毒序列, 其中314条比对长度超过1 000 nt, 54条覆盖病毒基因组90%以上的近全长序列。

表3 RNA-seq数据组装得到的病毒结果汇总
Table 3 Count of virus assembled from RNA-seq data

项目 Item		DM	CS	TU	TD	PV	PG	LBW	LBT	CR	Total
PVH	Total virus contig	24	0	0	0	67	29	0	0	111	231
	Contig of BLAST length > 1 000 nt	10	0	0	0	4	11	0	0	3	28
	Contig cover 90% of virus genome	2	0	0	0	0	2	0	0	0	4
PLRV	Total virus contig	0	0	0	0	4	0	1	1	1	7
	Contig of BLAST length > 1 000 nt	0	0	0	0	3	0	0	0	0	3
	Contig cover 90% of virus genome	0	0	0	0	0	0	0	0	0	0
PVM	Total virus contig	409	7	0	0	0	2	0	0	245	663
	Contig of BLAST length > 1 000 nt	21	0	0	0	0	0	0	0	33	54
	Contig cover 90% of virus genome	2	0	0	0	0	0	0	0	3	5
PVS	Total virus contig	54	0	26	29	252	23	0	0	412	796
	Contig of BLAST length > 1 000 nt	12	0	5	6	9	7	0	0	110	149
	Contig cover 90% of virus genome	4	0	0	5	0	6	0	0	13	28
PVX	Total virus contig	0	0	2	0	0	109	12	0	4	127
	Contig of BLAST length > 1 000 nt	0	0	1	0	0	10	0	0	0	11
	Contig cover 90% of virus genome	0	0	0	0	0	1	0	0	0	1
PVY	Total virus contig	66	0	1	0	10	12	0	0	23	112
	Contig of BLAST length > 1 000 nt	18	0	1	0	3	4	0	0	0	26
	Contig cover 90% of virus genome	4	0	0	0	2	1	0	0	0	7
ToCV-RNA1	Total virus contig	0	0	0	0	0	80	0	0	0	80
	Contig of BLAST length > 1 000 nt	0	0	0	0	0	8	0	0	0	8
	Contig cover 90% of virus genome	0	0	0	0	0	0	0	0	0	0
ToCV-RNA2	Total virus contig	0	3	0	0	0	48	0	0	0	51
	Contig of BLAST length > 1 000 nt	0	1	0	0	0	21	0	0	0	22
	Contig cover 90% of virus genome	0	0	0	0	0	3	0	0	0	3
TuMV	Total virus contig	0	0	0	0	0	0	0	0	21	21
	Contig of BLAST length > 1 000 nt	0	0	0	0	0	0	0	0	2	2
	Contig cover 90% of virus genome	0	0	0	0	0	0	0	0	1	1
CYVCV	Total virus contig	0	0	0	0	0	15	0	0	0	15
	Contig of BLAST length > 1 000 nt	0	0	0	0	0	9	0	0	0	9
	Contig cover 90% of virus genome	0	0	0	0	0	5	0	0	0	5
MYMV-A	Total virus contig	0	0	0	0	0	0	12	0	0	12
	Contig of BLAST length > 1 000 nt	0	0	0	0	0	0	0	0	0	0
	Contig cover 90% of virus genome	0	0	0	0	0	0	0	0	0	0
MYMV-B	Total virus contig	0	0	0	0	0	0	14	0	0	14
	Contig of BLAST length > 1 000 nt	0	0	0	0	0	0	2	0	0	2
	Contig cover 90% of virus genome	0	0	0	0	0	0	0	0	0	0

组装得到的病毒中PLRV、PVH、PVM、PVS、PVX、PVY、番茄褪绿病毒(Tomato chlorosis virus, ToCV)已被报道对马铃薯有侵染能力。香石竹潜隐病毒属(*Carlavirus*)3个成员PVS、PVM和PVH序列最多, 占全部序列的79.38%。其中以PVS序列为主, 9个数据样本中有6个组装得到PVS序列, 共796条序列, 包括149条比对长度大于1 000 nt的序列, 并组装得到28条近全长序列(达到对应病毒全长序列的90%以上), 长度范围为7 673~8 475 nt, 占PVS全长(8 478 nt)的90.50%~99.96%。PVM数量仅次于PVS, 共663条序列, 包括54条比对长度大于1 000 nt的序列和5条近全长序列, 长度范围为8 254~8 522 nt, 占PVM全长(8 533 nt)的96.73%~99.64%。同时, 从DM、PG、PV和CR样品中组装得到PVH序列231条, 包括28条比对长度大于1 000 nt的序列和4条PVH近全长序列, 长度范围为7 989~8 407 nt, 占PVH全长(8 410 nt)的94.99%~99.96%, 目前NCBI数据库中仅有141条PVS全长序列和7条PVH全长序列, 此结果丰富了PVS和PVH数据库的核酸序列资源。PVX和PVY都组装得到100条左右的序列, 包括1条PVX近全长序列, 长度为6 434 nt, 占PVX全长(6 435 nt)的99.98%, 7条PVY近全长序列, 长度范围为8 915~9 691 nt, 占PVY全长(9 704 nt)的91.87%~99.87%。PLRV组装结果较少, 仅有7条序列, 3条大于1 000 nt, 未获得近全长序列。ToCV有RNA1和RNA2两部分在PG样本中均被检测到, 而且组装得到3条RNA2近全长序列, 长度范围为7 539~8 221 nt, 占RNA2全长(8 247 nt)的91.42%~99.68%。

除组装得到上述7种已报道的对马铃薯有侵染能力的病毒之外, 本试验还在组装结果中发现芜菁花叶病毒(Turnip mosaic virus, TuMV)、柑橘黄化脉明病毒(Citrus yellow vein clearing virus, CYVCV)和绿豆黄花叶病毒(Mungbean yellow mosaic virus, MYMV)3种尚未报道对马铃薯有侵染能力的病毒。本试验组装到21条TuMV序列, 有2条比对长度大于1 000 nt, 其中1条近全长, 与NCBI数据库中TuMV的BJ-B04分离物(KC119188.1)

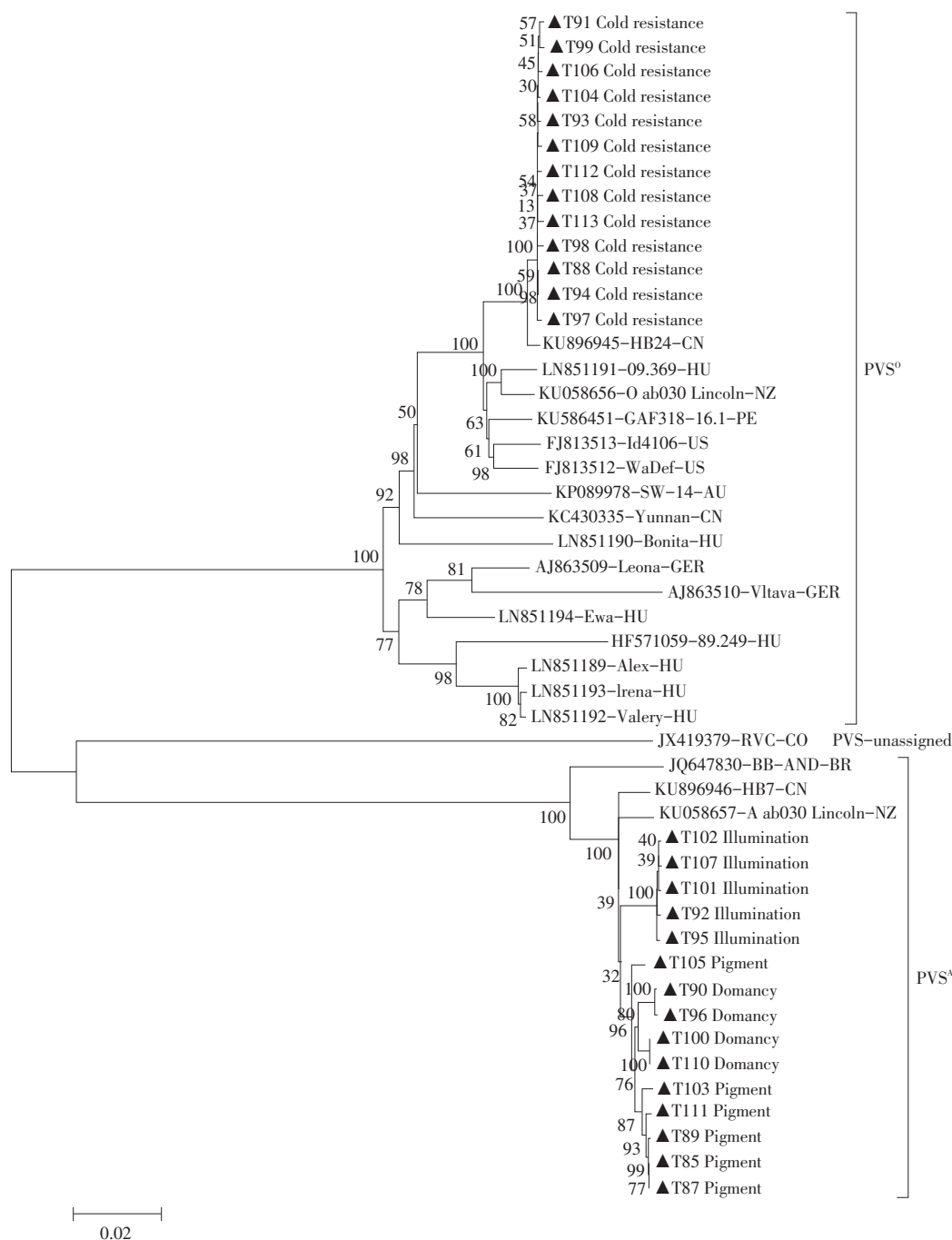
的同源性高达96.25%。组装得到15条CYVCV序列, 9条比对长度大于1 000 nt, 5条近全长序列, 与NCBI数据库中CYVCV的CQ分离物(KP313240.1)同源性达99%。本试验同时组装得到MYMV病毒的DNA-A和DNA-B两部分序列, 其中2条1 000 nt以上的MYMV病毒DNA-B序列已占到基因组40%以上, 其与NCBI库中MYMV的MF1分离物的DNA-B(JQ398670.1)同源性达98%。因此推测马铃薯可能是TuMV、MYMV、CYVCV 3种病毒的潜在寄主。

2.3 PVS病毒序列系统进化分析

本试验重点对组装得到的28条PVS近全长序列, 与NCBI数据库中20条PVS全长序列进行系统进化分析。全部48个PVS分离物除了PVS RVC-CO外, 主要分为PVS⁰和PVS^A两类。其中CR来源的序列聚类于PVS⁰, 与中国的PVS分离物HB24亲缘关系最近。PG、DM、TD来源的序列聚类于PVS^A, 与新西兰的NZ-A ab030 Lincoln分离物以及中国的HB7分离物亲缘关系最近(图1)。该结果表明这些样品中同时存在PVS^A和PVS⁰两种株系类型, 其中未发现被PVS^A和PVS⁰同时侵染的植物单株, 也未发现两者的重组株系。

2.4 RT-PCR验证测序数据组装结果

为验证测序数据分析结果, 本试验以9个转录组数据的原始RNA样本为模板, 对PVH、PVM、PVS、PVX、PVY和PLRV 6种病毒进行RT-PCR检测, 并将检测结果与NGS组装结果进行对比, 验证组装结果准确性。结果显示, 在DM、PV、PG和CR 4个样本中检测到PVH(图2A), 在DM、CS、PG和CR 4个样本中检测到PVM(图2B), 在DM、TU、TD、PV、PG和CR 6个样本中检测到PVS(图2C), 在TU、PG、LBW和CR 4个样本中检测到PVX(图2D), 在DM、TU、PV、PG和CR 5个样本中检测到PVY(图2E), RT-PCR验证结果与对测序数据的分析结果完全一致(表3)。此外, 在DM、CS、TD、PV、LBW、LBT和CR 7个样本中都检测到PLRV(图2F), 但测序数据的组装分析, 仅在PV、LBW、LBT、CR 4个样本中检测到极少的PLRV相关序列(表3), RT-PCR验证结果与测序分析结果差异较大。

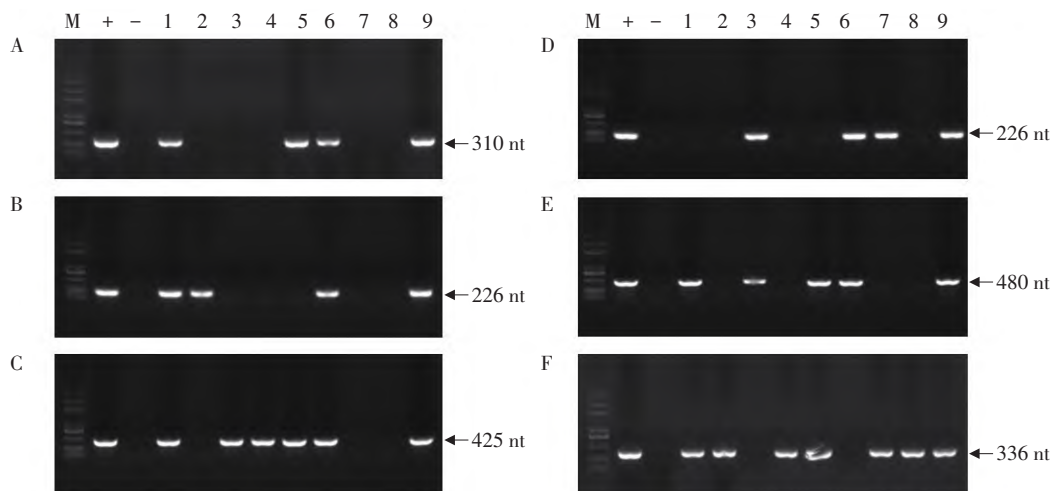


注: NCBI数据库中各分离物用GenBank登录号、分离物名称以及国籍表示, 组装得到的PVS序列用数据来源样本加给定的ID命名, 并用▲标注。标尺表示遗传距离, 节点上的数值表示置信度。

Note: Each isolate from the NCBI database is represented by the GenBank number, isolate name, and nationality. The assembled PVS sequence are named with the data source samples and the given IDs, and labeled with ▲. The scale represents the genetic distance, and the numerical value on the node represents the confidence level.

图1 组装得到的28个PVS基因组与NCBI中20个PVS全长基因组的系统进化分析

Figure 1 Phylogenetic tree of 28 assembled PVS genome and 20 complete PVS genome from NCBI



注: M代表Marker(Trans 2K-plus), 从上到下分别代表5 000、3 000、2 000、1 000、750、500、250和100 nt; +代表阳性对照, -代表阴性对照; 泳道1~9分别代表DM、CS、TU、TD、PV、PG、LBW、LBT和CR的原始RNA样本。A~F分别代表PVH、PVM、PVS、PVX、PVY、PLRV 6种病毒。

Note: M stands for Marker (Trans 2K-plus), representing 5 000, 3 000, 2 000, 1 000, 750, 500, 250 and 100 nt from top to bottom, respectively. "+" represents positive control; "-" represents negative control. Lanes 1~9 represent the original RNA samples of DM, CS, TU, TD, PV, PG, LBW, LBT and CR, respectively. A~F represents PVH, PVM, PVS, PVX, PVY and PLRV, respectively.

图2 转录组原始RNA样品6种病毒RT-PCR检测结果

Figure 2 Result of viral RT-PCR detection on the original RNA sample of the transcriptome

3 讨论

本试验收集到的RNA-seq数据中20%~30% Reads无法匹配马铃薯参考基因组, 可能有以下2个原因: (1)测序材料与马铃薯参考基因组测序所用材料‘DM1-3’(*Solanum phureja*)差异较大, ‘DM1-3’属于原始栽培种, 而测序所用材料大部分是普通栽培种或者野生种, 如‘Eshu3’(*S. tuberosum*)、‘ED25’(*S. tuberosum*)属于普通栽培种, ‘40-3’(*S. chacoense*)属于野生种, 种间差异导致部分Reads无法匹配; (2)测序材料可能存在病毒侵染的情况, 如DM样本中有31.24% Unmapped reads, 其中20.24%来源于病毒, 占Unmapped reads的60%以上。RNA-seq数据中病毒数据占1.79%, 组装得到10种病毒序列, 包括9种RNA病毒和1种DNA病毒, 其中PLRV、PVH、PVM、PVS、PVX、PVY和ToCV已被报道能侵染马铃薯^[15], 而组装得到的TuMV、CYVCV、MYMV是本试验首次发现对马铃薯有侵染能力的病毒。对组

装得到的28条PVS序列的系统进化分析表明, 中国侵染马铃薯的PVS包括PVS^A和PVS⁰两种株系类型, 与Wang等^[16]的试验结果吻合, 但本试验并未发现被PVS^A和PVS⁰同时侵染的植物单株, 无法确定是否存在这两种株系类型的天然复合侵染, 也未发现两者的重组株系, 值得进一步探究。通过RT-PCR验证结果表明, PVM、PVS、PVX、PVY和PVH的RT-PCR验证结果与NGS组装结果一致, 但在多个样品中可通过RT-PCR检测到PLRV, 测序分析无法检测到, 其原因可能是这些转录本测序数据深度和质量低。因为转录组测序建库时主要有两种方式富集mRNA: 去核糖体RNA富集mRNA和通过Oligo(dT)法富集有Poly(A)尾巴的mRNA, 本试验收集的数据大多采用后一种建库方式。因此, 没有Poly(A)尾巴的PLRV相对其他含Poly(A)尾巴的病毒占比偏低, 在测序数据质量和深度低的情况下, 难以组装得到PLRV序列。

相比传统病毒检测方法, 如电子显微镜检

测、指示植物接种、芯片杂交、血清学检测方法以及分子生物学检测法等^[17], 高通量测序技术具有明显优势。(1)快速、灵敏。高通量测序技术一般仅需2~3 d即可测完一个反应, 即使病毒滴度较低, 样本依然可使用高通量测序技术进行检测。(2)广谱性。高通量测序技术能同时对多个样本中的多种病毒进行全面检测鉴定, 包括DNA病毒、RNA病毒和类病毒, 均能被检测到。(3)不需预先知道病毒背景信息。可加快新病毒的发现历程, 这是传统病毒检测研究手段无法实现的^[18]。本试验收集接近1 T的RNA-seq数据进行组装, 组装得到10种病毒的54条近全长序, 虽然材料中病毒滴度较高, 但仍难组装得到完整的病毒基因组序列, 因软件组装过程中会产生Gap, 所以很难组装得到完整病毒基因组序列, 需根据组装得到的病毒序列借助RACE技术进行扩增。

随着设备更新和技术进步, 三代测序技术登上舞台, 其读长更长, 更适合组装, 会为植物病毒研究带来新机遇。

[参 考 文 献]

- [1] 谢从华, 柳俊. 中国马铃薯科技发展与创新之回顾 [J]. 华中农业大学学报, 2021, 40(4): 16–26.
- [2] 吴兴泉, 时妍, 杨庆东. 我国马铃薯病毒的种类及脱毒薯生产过程中病毒的检测 [J]. 中国马铃薯, 2011, 25(6): 363–366.
- [3] 邱彩玲, 韩树新, 白艳菊, 等. 马铃薯生产中马铃薯纺锤块茎类病毒的综合防治 [J]. 中国马铃薯, 2017, 31(3): 154–159.
- [4] Wang Q, Zhang W. An economic analysis of potato demand in China [J]. American Journal of Potato Research, 2010, 87: 245–252.
- [5] Xu X, Pan S, Cheng S, *et al.* Genome sequence and analysis of the tuber crop potato [J]. Nature, 2011, 475(7355): 189–195.
- [6] Pham G M, Hamilton J P, Wood J C, *et al.* Construction of a chromosome-scale long-read reference genome assembly for potato [J]. Gigascience, 2020, 9(9): giaa100.
- [7] Yang X, Zhang L, Guo X, *et al.* The gap-free potato genome assembly reveals large tandem gene clusters of agronomical importance in highly repeated genomic regions [J]. Molecular Plant, 2023, 16(2): 314–317.
- [8] Batty E M, Wong T H N, Trebes A, *et al.* A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples [J]. PloS one, 2013, 8(6): e66129.
- [9] Sparks M E, Gundersenrindal D E, Harrison R L. Complete genome sequence of a novel iflavirus from the transcriptome of *Halyomorpha halys*, the brown marmorated stink bug [J]. Genome Announcements, 2013, 1(6): e00910–13.
- [10] Smith G, Macias-Muñoz A, Briscoe A D. Genome sequence of a novel iflavirus from mRNA sequencing of the butterfly *Heliconius erato* [J]. Genome Announcements, 2014, 2(3): e00398–14.
- [11] Sun X X, Niu J X, Wang B H, *et al.* Pear is a new host of hop stunt viroid [J]. Basic and Clinical Pharmacology and Toxicology, 2015, 117 (Suppl 4): 23.
- [12] 李伟佳, 张志宏, 郭巍, 等. 利用高通量测序技术快速解析草莓病毒基因组序列 [J]. 沈阳农业大学学报, 2016, 47(5): 536–540.
- [13] 胡国君, 董雅凤, 张尊平, 等. 用于苹果树病毒鉴定的 small RNA 深度测序数据集 [J]. 农业大数据学报, 2022, 4(2): 25–29.
- [14] 张双双. 马铃薯表达谱平台的构建与应用 [D]. 武汉: 华中农业大学, 2015.
- [15] Kreuze J F, Souza-Dias J A C, Jeevalatha A, *et al.* Viral diseases in potato [M]//Campos H, Ortiz O. The potato crop: Its agricultural, nutritional and social contribution to humankind. Cham, Switzerland: Springer, 2020: 389–430.
- [16] Wang J, Meng F, Chen R, *et al.* RT-PCR differentiation, molecular and pathological characterization of Andean and ordinary strains of potato virus S in potatoes in China [J]. Plant Disease, 2016, 100(8): 1580–1585.
- [17] 刘玲玲, 韩黎明, 张尚智, 等. 马铃薯病毒的常用检测方法 [J]. 中国马铃薯, 2013, 27(4): 239–242.
- [18] 钱亚娟, 徐毅, 周琦, 等. 利用深度测序技术发掘植物病毒资源 [J]. 中国科学: 生命科学, 2014, 44(4): 351–363.