

黄皮叶绿体基因组测序与分析

赵志常^{1,2}, 高爱平^{1,2*}, 黄建峰^{1,2}, 罗睿雄^{1,2}

(1. 中国热带农业科学院热带作物品种资源研究所/农业部华南作物基因资源与种质创制重点开发实验室, 海南海口 571101; 2. 国家热带果树品种改良中心, 海南海口 571101)

摘要 [目的] 分析黄皮 [*Clausena lansium* (Lour.) Skeels] 叶绿体基因组, 为黄皮属不同种或品种的进化、杂交、演变, 以及黄皮不同品种的鉴定等提供技术支持。[方法] 使用试剂盒提取了黄皮的叶绿体 DNA, 通过测序、组装、注释获得了黄皮叶绿体基因组。[结果] 黄皮叶绿体基因组全长 159 283 bp, 其中反相重复序列区 (IRs) 长 53 998 bp, 大单拷贝序列区 (LSC) 和小单拷贝序列区 (SSC) 长度分别为 87 301、17 983 bp, 共注释 126 个基因, 包括编码蛋白基因 89 个, tRNA 基因 29 个和 rRNA 基因 8 个。黄皮叶绿体基因组全序列 GC 含量 38.7%。对 5 种芸香科果树的全长序列、SSC、LSC、IRA、IRB 进行比较分析发现, 假黄皮的叶绿体全长序列、LSC、IRA、IRB 区域的长度较其他 4 种果树长, 柠檬的 SSC 区域长度最长, 黄皮的全长序列和 SSC 区域长度最短。黄皮叶绿体的基因数目和编码蛋白数目较其他 4 种果树多。对 20 种园艺植物的叶绿体基因组进行分析发现, 同一科下面的不同属植物或同一属下面的不同种更容易聚为一类。[结论] 该研究丰富了热带果树的叶绿体基因组数据库, 为种质资源的合理开发利用提供了依据。

关键词 黄皮; 叶绿体基因组; 测序与分析

中图分类号 S667.9 文献标识码 A

文章编号 0517-6611(2019)11-0115-04

doi: 10.3969/j.issn.0517-6611.2019.11.032



开放科学(资源服务)标识码(OSID):

Sequencing and Analysis of Chloroplast Genome of *Clausena lansium* (Lour.) Skeels

ZHAO Zhi-chang^{1,2}, GAO Ai-ping^{1,2}, HUANG Jian-feng^{1,2} et al (1. Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences/Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Haikou, Hainan 571101; 2. National Cultivar Improvement Center of Tropical Fruit Tree, Haikou, Hainan 571101)

Abstract [Objective] The analysis of the chloroplast genome of *Clausena lansium* (Lour.) Skeels provides technical support for the evolution, hybridization and evolution of different species or varieties of *Clausena lansium*, as well as the identification of different varieties of *Clausena lansium*. [Method] The chloroplast DNA was extracted from *Clausena lansium* by kit, and chloroplast genome was obtained by sequencing, assembly and annotation. [Result] The chloroplast genome is 159 283 bp in length, of which the inverted repeats sequence (IRs) is 53 998 bp, the large single-copy region (LSC) and the small single copy region (SSC) are 87 301 and 17 983 bp, respectively. A total of 126 genes are annotated, including 89 coding protein genes, 29 tRNA genes and 8 rRNA genes. The full-sequence GC content of the chloroplast genome was 38.7%. The full-length sequences, SSC, LSC, IRA and IRB of five Rutaceae fruit trees, were compared and analyzed. It was found that the chloroplast full-length sequences, LSC, IRA and IRB regions of *Clausena excavata* were longer than those of the other four fruit trees, and the *Citrus limon* in SSC region was the longest and the *Clausena lansium* was the shortest. And the whole length sequence of *Clausena lansium* is also the shortest among the five species. The number of genes and coding proteins of chloroplasts in *Clausena lansium* was more than that in other four fruit trees. Analysis of the chloroplast genomes of 20 horticultural plants revealed that different genus plants under the same family or different species under the same genus were more likely to be clustered into one group. [Conclusion] This study enriched the chloroplast genome database of tropical fruit trees and provided a basis for the rational development and utilization of germplasm resources.

Key words *Clausena lansium* (Lour.) Skeels; Chloroplast genome; Sequencing and analysis

叶绿体是光合作用的重要细胞器, 探讨叶绿体各部分的功能以及在生物进化中的系统关系及作用, 对研究作物改良、提高光合效率、探索细胞器的起源和进化具有重要意义。全面分析叶绿体基因组序列是做好这些工作的前提。叶绿体基因组研究不仅有助于通过遗传转化体系改良叶绿体功能, 选育出有价值的新品种, 而且可以在分子水平上揭示光合作用机理、核质互作关系以及物种起源与进化。1986年 Shi-nozaki 等^[1] 获得烟草叶绿体基因组全序列, 开启了叶绿体基因组测序的先河。目前, 叶绿体基因组数据库正在迅速充实, 到 2018 年 1 月 NCBI 中已收集了来自不同植物的 4 000 多条叶绿体基因组全序列, 以及一些果树植物的部分叶绿体基因组或某些基因的序列, 其中包含同一个种的不同亚种和同一个种的不同栽培品种的叶绿体基因组全序列。

叶绿体基因组非常保守, 主要表现在基因组结构、基因排列顺序和基因种类等方面。在高等植物中, 叶绿体 DNA 以共价、闭合、环形双链 DNA 的多拷贝形式存在, 大小一般在 120~160 kb, 因此具有单亲遗传, 小分子量、多拷贝、分子进化速率慢等特点^[2-4]。叶绿体 DNA 序列已被广泛应用于分子系统发育研究中。该研究以黄皮叶绿体基因组原始测序数据结果为基础, 进一步对黄皮叶绿体全基因组序列进行注释、校正, 再将其与芸香科的 4 种植物叶绿体基因组进行比较, 并对 20 种园艺植物的叶绿体基因序列进行聚类分析, 以期对黄皮不同品种的鉴定以及种质资源的开发利用提供依据。

1 材料与方法

1.1 材料 新鲜的黄皮叶片取中国热带农业科学院热带作物品种资源研究所黄皮种质资源圃, -20 °C 冰箱保存备用。

1.2 研究方法

1.2.1 黄皮叶绿体基因组 DNA 的提取与组装。 利用天根公司的植物 DNA 提取试剂盒提取黄皮叶片的总 DNA。DNA 构建 300 bp 的文库, 在 Illumina 测序平台, 总测序深度约为 129X, 测序得到约 39 G 的数据, 对原始数据进行过滤低质量

基金项目 国家自然科学基金资助项目(31471850); 国家重点研发计划(2018YFD1000500); 非营利性科研机构改革启动经费(1630032017005; 1630032017004)。

作者简介 赵志常(1977—), 男, 山东烟台人, 副研究员, 从事热带果树遗传育种与分子生物学研究。* 通信作者, 研究员, 从事果树资源学方面研究。

收稿日期 2019-01-14

和 PCR 复制等处理后,得到高质量的数据,使用 SOAPdenovo 软件对高质量数据进行组装,并以假黄皮的序列作为参考。采用 OGDRAW 软件(<http://ogdraw.mpimp-golm.mpg.de/>)生成叶绿体基因组的物理图谱^[5]。

1.2.2 叶绿体基因组的特征分析及系统进化分析。对照 NCBI 上已经登录甜橙(DQ864733)、假黄皮(NC_032685)、柠檬(KJ865401)、山小橘(KU949005)4种植物的叶绿体基因组序列,进行5种果树的叶绿体基因组的比较。根据 NCBI 上已经登录植物的叶绿体基因组序列共选取了20个主要园艺植物的叶绿体基因组序列,利用 BioEdit 软件对各物种的10 000个核苷酸序列利用近邻系统发育树分析^[6-7]。

2 结果与分析

2.1 黄皮叶绿体全基因组结构及特征 结果表明,黄皮叶

绿体基因组呈环状双链。与大多数高等植物叶绿体基因组一样,存在2个反向重复序列(Inverted repeat, IR),即 IRA 和 IRB;反向重复序列之间有一个大单拷贝区(Large single-copy region, LSC)和一个小单拷贝区(Small single-copy region, SSC)。黄皮叶绿体基因组全长159 283 bp,其中反相重复序列(IRs)长53 998 bp,大单拷贝序列区(LSC)和小单拷贝序列区(SSC)长度分别为87 301、17 983 bp,共注释126个基因,包括编码蛋白基因89个,tRNA 基因29个和rRNA 基因8个(图1)。在全基因组中,A+T 含量为61.4%、G+C 含量为38.6%,在SSC中A+T 含量为66.6%、G+C 含量为33.4%,LSC中A+T 含量为62.9%、G+C 含量为37.1%,IRA 中A+T 含量为57.0%、G+C 含量为43.0%,IRB 中A+T 含量为57.1%、G+C 含量为42.9%(表1)。

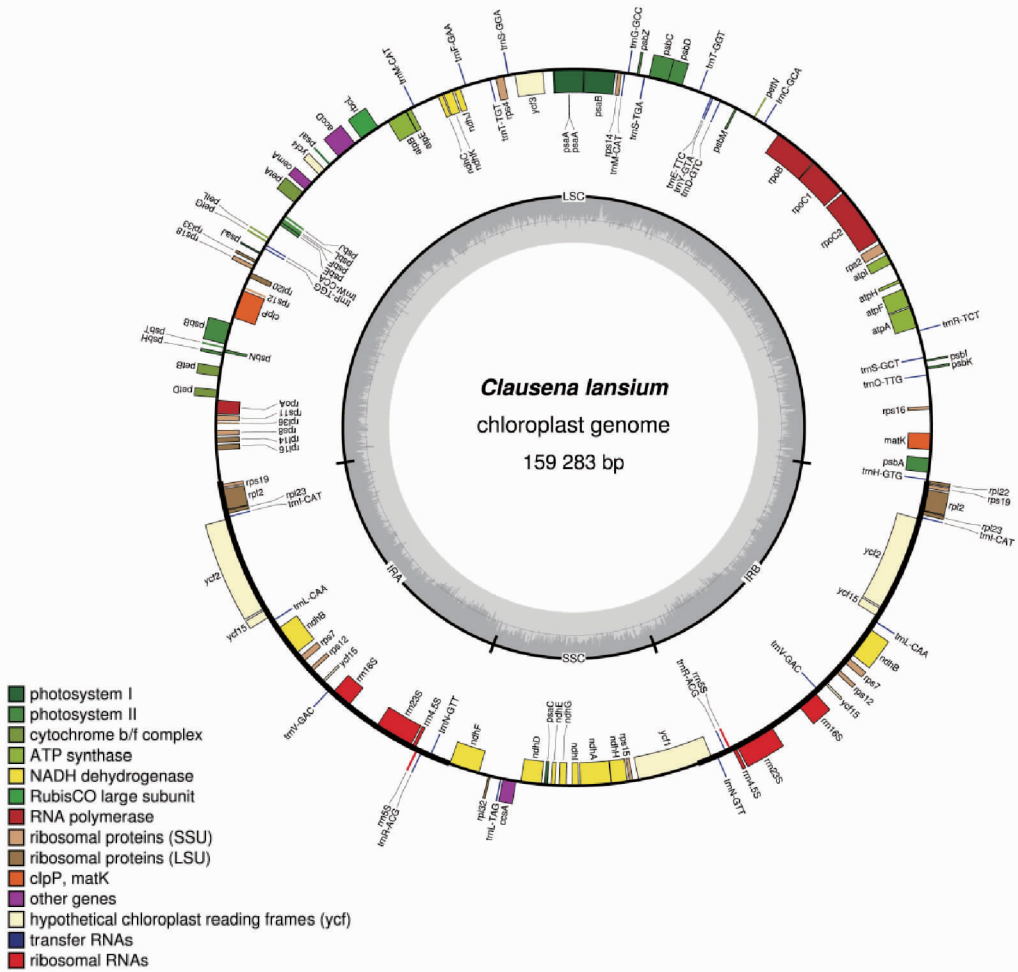


图1 黄皮叶绿体基因组的物理图谱

Fig.1 The physical map of chloroplast genome in *Chlausena lansium*

2.2 5种芸香科叶绿体基因组比较 对 NCBI 上已经登录的假黄皮叶绿体基因组序列进行分析发现,其全长161 171 bp,反向重复区(IRA 和 IRB, 27 411bp)、小单拷贝区(SSC, 18 029 bp)、大单拷贝区(LSC, 88 055 bp),其叶绿体基因组有113个基因,80个编码蛋白质基因,30个转运 RNA 基因和4个核糖体 RNA 基因。甜橙叶绿体全基因组长160 129 bp,大单拷贝区(LSC, 87 744 bp)、小单拷贝区(SSC, 18 393 bp)、反向重复区(IRA 和 IRB, 26 991 bp),其叶绿体基因组有113

个基因、79个编码蛋白质基因、30个转运 RNA 基因和4个核糖体 RNA 基因。柠檬叶绿体全基因组长159 893 bp,大单拷贝区(LSC, 87 148 bp)、小单拷贝区(SSC, 18 763 bp)、反向重复区(IRA 和 IRB, 26 991 bp),其叶绿体基因组有115个基因、81个编码蛋白质基因、30个转运 RNA 基因和4个核糖体 RNA 基因。山小橘叶绿体全基因组长159 844 bp,大单拷贝区(LSC, 87 494 bp)、小单拷贝区(SSC, 18 332 bp)、反向重复区(IRA 和 IRB, 27 009 bp),其叶绿体基因组有117个基

因、83 个编码蛋白质基因、30 个转运 RNA 基因和 4 个核糖体 RNA 基因。对 5 种热带果树的全长序列、SSC、LSC、IRA、IRB 进行比较分析发现,假黄皮的叶绿体全长序列、LSC、IRA、IRB 区域的长度较其他 4 种果树长,SSC 区柠檬的最

长,黄皮的最短。并且黄皮的全长序列也是 5 种植物中最短的。黄皮叶绿体的基因数目和编码蛋白数目较其他 4 种果树多。作为芸香科的假黄皮和黄皮两者在 tRNA 和 rRNA 的数目上均有差异。

表 1 黄皮叶绿体基因组中各种碱基的百分比含量

Table 1 Percentage contents of various bases in chloroplast genome of *Clausena lansium*

区域 Region	碱基百分比含量 Percentage contents of various bases//%					
	A	T	C	G	A+T	C+G
叶绿体基因组 Chloroplast genome	30.4	31.0	19.6	19.0	61.4	38.6
小单拷贝区 Small single-copy region, SSC	33.4	33.2	17.2	16.2	66.6	33.4
大单拷贝区 Large single-copy region, LSC	30.9	32.0	19.1	18	62.9	37.1
反向重复区 A A inverted repeated regions A, IRA	28.6	28.4	20.8	22.2	57.0	43.0
反向重复区 B A inverted repeated regions B, IRB	28.5	28.6	22.2	20.7	57.1	42.9

表 2 5 种芸香科植物叶绿体基因组的特征分析

Table 2 Comparisons among the chloroplast genome characteristics of five Rutaceae plants

物种 Species	基因组大小 Genome size	大单拷贝区 Large single-copy region, LSC	小单拷贝区 Small single-copy region, SSC	反向重复区 Inverted repeated regions A/B, IRA/B	基因数目 Number of protein	编码蛋白数目 Number of protein	tRNAs/rRNAs 的数目 Number of tRNAs/rRNAs
黄皮 <i>Clausena lansium</i>	159 283	87 301	17 984	26 999	126	89	29/8
假黄皮 <i>Clausena excavata</i>	161 172	88 055	18 029	27 411	113	80	30/4
甜橙 <i>Citrus sinensis</i>	160 129	87 744	18 393	26 996	113	79	30/4
柠檬 <i>Citrus limon</i>	159 893	87 148	18 763	26 991	115	81	30/4
山小橘 <i>Glycosmis pentaphylla</i>	159 844	87 494	18 332	27 009	117	83	30/4

2.3 叶绿体基因组系统进化分析 根据 NCBI 登录的 20 种园艺植物的叶绿体基因组序列进行聚类分析发现,芸香科柑橘属 (*Citrus*) 的酸柑 (*Citrus aurantiifolia*)、拜伦基尔 (*Citrus platyamma*)、脐橙 (*Citrus sinensis*)、山小橘属 (*Glycosmis*) 的毛糖糖 (*Glycosmis mauritiana*)、山小橘 (*Glycosmis pentaphylla*)、九里香属 (*Murraya*) 的九里香 (*Murraya koenigii*) 和黄皮属 (*Clausena*) 的假黄皮 (*Clausena excavata*) 等植物的叶绿体可以聚在一起,表明这些属植物之间的叶绿体相似性比较高。梨属 (*Pyrus*) 的沙梨 (*Pyrus pyrifolia*)、柊木属 (*Docynia*) 的多依果 (*Docynia delavayi*)、榲桲属 (*Cydonia*) 的榲桲 (*Cydonia oblonga*)、枇杷属 (*Eriobotrya*) 的枇杷 (*Eriobotrya japonica*) 都属于蔷薇科植物 (*Rosaceae*) 可以聚在一起。费约果 (*Acca sellowiana*)、番石榴 (*Psidium guajava*)、嘉宝果 (*Plinia trunciflora*)、海南蒲桃 (*Syzygium cumini*) 等都属于桃金娘科 (*Myrtaceae*) 植物可以聚在一起 (图 3)。荔枝 (*Litchi chinensis*) 属于无患子科在叶绿体上与其他植物的关系较远,而黄皮叶绿体基因组与芸香科植物的叶绿体有一定的差异。黄皮和荔枝都是原产于我国的热带、亚热带果树,两者的叶绿体基因组序列与其他物种有较大的差异,这可能由于长期的人为选择或热带、亚热带的环境造成的。

3 讨论

大多数植物叶绿体基因组长度一般为 107~218 kb,对园艺植物而言,其叶绿体一般分布在 150~170 kb 较多,如甜橙的叶绿体为 160 kb^[8]。草莓属的叶绿体基因组一般在 155 kb^[9],葡萄属的植物叶绿体基因组一般在 160 kb^[10]。而同为芸香科的假黄皮和黄皮两者的叶绿体存在一定的差异,前者 161 kb 而后者 159 kb 并且在 SSC、LSC、IRA、IRB 的长度

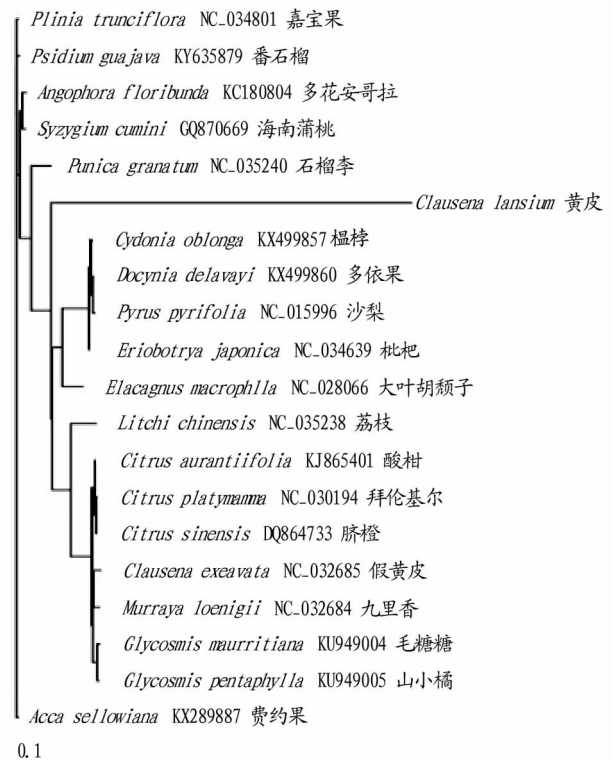


图 3 20 种园艺植物的叶绿体基因组的聚类分析

Fig. 3 Cluster analysis of chloroplast genomes of 20 horticultural plants

上均是后者小于前者。两者在叶片的大小和形态上也有一定的差异,前者的叶片较小、叶片浅绿色,而后者的叶片较大,叶片一般呈绿色,在果实大小和形态方面也有一定的差异。叶绿体基因组结构比较保守,但基因组大小却不尽相同,

叶绿体基因组差异主要是由反向重复区的收缩、扩张或缺失引起的,而 IR 在稳定叶绿体基因组结构和影响叶绿体基因组大小方面起着非常重要的作用^[11-12]。5 种芸香科植物的 IR 区存在一定的差异,作为水果食用的甜橙、柠檬、黄皮其 IR 区域在 26 991~26 999 bp,而作为一种野生果树或观赏果树的假黄皮和山小橘两者的 IR 区域在 27 kb 左右,同时,作为观赏果树的九里香(*Murraya koenigii*)其 IR 区域也在 27 kb。IR 区是不是与芸香科果实的可食性有一定关系有待进一步研究。

近年来,植物叶绿体基因组的研究得到国内外众多学者的普遍关注,现已成为叶绿体转化技术和植物亲缘关系鉴定的一个重要手段。通过对 20 种园艺植物的叶绿体基因组分析发现,亲缘关系较近的同属植物、同属不同种的植物更容易聚在一起。笔者首次采用高通量测序技术完成了黄皮叶绿体基因组序列、结构和注释工作,丰富了黄皮的遗传资源,对 5 种芸香科植物的叶绿体基因组的信息进行了比较分析,为热带果树的叶绿体研究提供了参考,丰富了热带果树的叶绿体基因组数据库,为种质资源的合理开发利用提供了基础性数据。

参考文献

- [1] SHINOZAKI K, OHME M, TANAKA M, et al. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression [J]. EMBO J, 1986, 5(9): 2043-2049.
- [2] WAKASUGI T, TSUDZUKI T, SUGIURA M. The genomics of land plant

- chloroplasts: Gene content and alteration of genomic information by RNA editing [J]. Photosynth Res, 2001, 70(1): 107-118.
- [3] KAHLAU S, ASPINALL S, GRAY J C, et al. Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes [J]. J Mol Evol, 2006, 63(2): 194-207.
- [4] SHIMADA H, SUGIURA M. Fine structural features of the chloroplast genome: Comparison of the sequenced chloroplast genome [J]. Nucleic Acids Res, 1991, 19(5): 983-995.
- [5] LOHSE M, DRECHSEL O, KAHLAU S, et al. Organellar Genome DRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. Nucleic acids research, 2013, 41: 575-581.
- [6] NEI M, KUMAR S. Molecular evolution and phylogenetics [M]. Oxford, USA: Oxford University Press, 2000.
- [7] YANG Z H, PAML Z. Phylogenetic analysis by maximum likelihood [J]. Mol Biol Evol, 2007, 24(8): 1586-1591.
- [8] BAUSHER M G, SINGH N D, LEE S B, et al. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': Organization and phylogenetic relationships to other angiosperms [J]. BMC Plant Biology, 2006, 6: 1-11.
- [9] BAI L J, YE Y T, CHEN Q, et al. The complete chloroplast genome sequence of the white strawberry *Fragaria pentaphylla* [J]. Conservation genetics resources, 2017, 9(4): 659-661.
- [10] XIE H K, JIAO J, FAN X C, et al. The complete chloroplast genome sequence of Chinese wild grape *Vitis amurensis* (Vitaceae: *Vitis* L.) [J]. Conservation genetics resources, 2017, 9(1): 43-46.
- [11] PALMER J D, THOMPSON W F. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost [J]. Cell, 1982, 29(2): 537-550.
- [12] WANG R J, CHENG C L, CHANG C C, et al. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots [J]. BMC Evolutionary Biology, 2008, 8(1): 36-51.

(上接第 114 页)

P80 的同源性为 97.3%。该试验结果表明,即使接种疫苗,仔猪依然存在 PRRSV 持续性感染状况。我国现有的商业化 PRRSV 弱毒疫苗能否有效预防 HP-PRRSV,尚有待进一步研究。

总而言之,该试验首次研究了全国规模化断奶仔猪携带 PRRSV 的现状,可为猪繁殖与呼吸综合征防控提供重要的数据和理论依据。

参考文献

- [1] NEUMANN E J, KLIEBENSTEIN J B, JOHNSON C D, et al. Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States [J]. J Am Vet Med Assoc, 2005, 227(3): 385-392.
- [2] 刘莹, 王凤雪, 温永俊, 等. 猪繁殖与呼吸综合征病毒非结构蛋白 2 的研究进展 [J]. 中国预防兽医学报, 2015, 37(2): 155-158.
- [3] HANADA K, SUZUKI Y, NAKANE T, et al. The origin and evolution of porcine reproductive and respiratory syndrome viruses [J]. Molecular biology & evolution, 2005, 22(4): 1024-1031.
- [4] MURTAUGH M P, STADEJEK T, ABRAHANTE J E, et al. The ever-expanding diversity of porcine reproductive and respiratory syndrome virus [J]. Virus research, 2010, 154(1): 18-30.
- [5] WANG X, MARTHALER D, ROVIRA A, et al. Emergence of a virulent porcine reproductive and respiratory syndrome virus in vaccinated herds in the United States [J]. Virus research, 2015, 210(1): 34-41.
- [6] LIU Y, WANG F X, WEN Y J, et al. Effect of nonstructural protein 2 hypervariable regions in the replication of porcine reproductive and respiratory syndrome virus in marc-145 cells [J]. Intervirology, 2015, 58(5): 288-296.
- [7] 付利芝, 徐登峰, 周淑兰, 等. 重庆地区猪繁殖与呼吸综合征病毒和猪瘟病毒感染检测 [J]. 中国兽医杂志, 2009, 45(4): 44-45.
- [8] TIAN K G, YU X L, ZHAO T Z, et al. Emergence of fatal PRRSV variants: Unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark [J]. PLoS One, 2007, 2(6): 1-10.

- [9] ZHAO K, YE C, CHANG X B, et al. Importation and recombination are responsible for the latest emergence of highly pathogenic porcine reproductive and respiratory syndrome virus in China [J]. Journal of virology, 2015, 89(20): 10712-10716.
- [10] ZHOU L, WANG Z C, DING Y P, et al. NADC30-like strain of porcine reproductive and respiratory syndrome virus, China [J]. Emerging infectious diseases, 2015, 21(12): 2256-2257.
- [11] GUO Z H, CHEN X X, LI R, et al. The prevalent status and genetic diversity of porcine reproductive and respiratory syndrome virus in China: A molecular epidemiological perspective [J]. Virology journal, 2018, 15(1): 1-14.
- [12] 祝永才, 胡兴义, 张双翔, 等. 复合益生菌对早期断奶仔猪生长性能和免疫水平的影响 [J]. 中国畜牧兽医, 2018, 45(6): 1518-1525.
- [13] HE B, LI Z S, YANG F L, et al. Virome profiling of bats from Myanmar by metagenomic analysis of tissue samples reveals more novel mammalian viruses [J]. PLoS One, 2013, 8(4): 1-14.
- [14] KANG H, YU J E, SHIN J E, et al. Geographic distribution and molecular analysis of porcine reproductive and respiratory syndrome viruses circulating in swine farms in the Republic of Korea between 2013 and 2016 [J]. BMC Veterinary Research, 2018, 14(1): 160.
- [15] SUI X K, XIN T, GUO X Y, et al. Genomic characterization and pathogenic study of two porcine reproductive and respiratory syndrome viruses with different virulence in Fujian, China [J]. Journal of veterinary science, 2018, 19(3): 339-349.
- [16] 徐晓杰, 张乔亚, 谈晨, 等. 2015-2016 年我国部分地区猪繁殖与呼吸综合征的分子流行病学调查 [J]. 畜牧与兽医, 2017, 49(6): 153-156.
- [17] ZHANG H, KONO H. Economic impacts of Porcine Reproductive and Respiratory Syndrome (PRRS) outbreak in Vietnam pig production [J]. Tropical agricultural research, 2012, 23(2): 152-159.
- [18] ZHU L Y, BI J L, ZHENG L L, et al. In vitro inhibition of porcine reproductive and respiratory syndrome virus replication by short antisense oligonucleotides with locked nucleic acid modification [J]. BMC Veterinary Research, 2018, 14(1): 109.
- [19] 赵耘, 罗长宝, 林志雄, 等. 猪繁殖和呼吸综合征病毒中国分离株 B₁₃ 株 ORF₅ 基因在杆状病毒系统中的表达 [J]. 农业生物技术学报, 2000, 8(3): 211-215.
- [20] 郭振华, 陈鑫鑫, 李睿, 等. 中国猪繁殖与呼吸综合征病毒流行历史及现状 [J]. 畜牧兽医学报, 2018, 49(1): 1-9.