

研究报告

丝状支原体山羊亚种 PG3 株的全基因组序列测定与分析

徐春光^{1,2} 郝永清^{2*} 郝瑞霞² 张玲² 刘波²

(1. 呼伦贝尔学院生命与环境科学学院 内蒙古 呼伦贝尔 021008)

(2. 内蒙古农业大学兽医学院 微生物学与免疫学实验室 内蒙古 呼和浩特 010018)

摘要:【目的】全面了解丝状支原体山羊亚种 PG3 株的全基因组序列信息, 寻找该病原体的主要保护性抗原基因。【方法】利用高通量 Illumina HiSeq 2000 测序技术对丝状支原体山羊亚种 PG3 株的全基因组进行测序与拼接, 借助软件和数据库对全基因组序列所承载的遗传信息进行注释和分析。【结果】丝状支原体山羊亚种 PG3 株基因组大小为 1 025 065 bp, G+C% 为 23.6%, 预测含有 846 个编码基因。根据 COG 分类和 KEGG 代谢通路分类, 基因组中绝大多数基因主要与蛋白翻译、核糖体结构与合成、DNA 复制与修复、糖代谢和环境信号传递与转换方面有关。该菌株具有丝状支原体山羊亚种特有的麦芽糊精/麦芽糖代谢途径。与 Mmc str. 95010 的基因组的比对结果显示二者具有良好的共线性关系。在基因组中发现 3 个串联排列的可变表面脂蛋白基因, 即 *GL000459*、*GL000461* 和 *GL000462*。【结论】获得丝状支原体山羊亚种 PG3 株的全基因组序列, 分析基因组基本特征, 初步解析 3 个串联排列的可变表面脂蛋白基因, 为进一步研究支原体可变表面脂蛋白的功能和研制生物工程疫苗奠定基础。

关键词: 丝状支原体山羊亚种, 全基因组序列测定, 可变表面脂蛋白

Complete genome sequence and analysis of *Mycoplasma mycoides* subsp. *capri* str. PG3

XU Chun-Guang^{1,2} HAO Yong-Qing^{2*} HAO Rui-Xia² ZHANG Ling² LIU Bo²

(1. Life & Environmental Sciences School, Hulunbuir College, Hulunbuir, Inner Mongolia 021008, China)

(2. Inner Mongolia Laboratory of Microbiology & Immunology, Veterinary School, Inner Mongolia Agriculture University, Hohhot, Inner Mongolia 010018, China)

Abstract: [Objective] To understand the complete genome sequence information of *Mycoplasma mycoides* subsp. *capri* str. PG3 (Mmc str. PG3) and offer help for screening of main antigenic genes. [Methods] The high flux Illumina HiSeq 2000 technology was used for genome sequencing and annotation of Mmc str. PG3. [Results] The genome of Mmc str. PG3 consisted of a circular chromosome of 1 025 065 bp with an overall molar G+C% content of only 23.6%. The chromosome of Mmc str. PG3 contained 846 predicted coding sequences. The majority of genes were associated with

Foundation item: Key Technologies R&D Program of China (No. 2011BAD18B01)

*Corresponding author: Tel: 86-471-4308677; E-mail: haoyq1960@163.com

Received: January 10, 2017; Accepted: March 27, 2017; Published online (www.cnki.net): March 27, 2017

基金项目: 国家科技支撑计划项目(No. 2011BAD18B01)

*通讯作者: Tel: 86-471-4308677; E-mail: haoyq1960@163.com

收稿日期: 2017-01-10; 接受日期: 2017-03-27; 优先数字出版日期(www.cnki.net): 2017-03-27

protein translation, ribosome synthesis and structure, DNA replication and repair, glycometabolism and environment transfer and conversion processes. The Mmc-specific maltodextrin/maltose gene cluster was also identified in the genome of Mmc str. PG3. Comparative genomic analysis showed that Mmc str. PG3 should have significant genomic synteny with Mmc str. 95010. Three genes (*GL000459*, *GL000461*, *GL000462*) were identified as variable lipoprotein genes in the genome of Mmc str. PG3 by genomic information analysis. **[Conclusion]** Complete genome sequence of Mmc str. PG3 provided in this study might facilitate genome-wide structural annotation to assess genes associated with pathogen virulence. The results of this study would provide a theoretical basis for further research on the role of variable surface lipoprotein of the pathogenesis of mycoplasma and development of biological engineering vaccine.

Keywords: *Mycoplasma mycoides* subsp. *capri*, Complete genome sequence, Variable lipoprotein

羊支原体性肺炎(Mycoplasmal pneumonia of sheep and goats, MPSG),又称为羊传染性胸膜肺炎,是由支原体引起的一种高度接触性传染病^[1-2]。在畜牧业以羊养殖业为主的不发达国家广泛流行,给各国带来了严重的经济损失^[3-5]。国内外研究者认为MPSG的病原复杂,存在多种致病菌型,而丝状支原体山羊亚种(*Mycoplasma mycoides* subsp. *capri*, Mmc)是广泛流行于我国很多地区 MPSG 疾病的优势菌型之一。Mmc 的菌体呈球状、棒状、弧状、逗号状、梨状等多种形态,以球状最为常见,比一般支原体大,直径为 150–1 250 nm。Mmc 营养要求比较苛刻,生长缓慢,在含有血清的固体培养基上形成透明的小菌落,比其他支原体菌落稍大,呈典型的油煎蛋状。PG3 株是 Mmc 的模式株,在 1951 年由 Chu 和 Berieidege 两人分离得到^[6]。由 Mmc 引起的 MPSG 疾病感染山羊,羔羊易感,一般不感染绵羊。该病在冬末春初和秋季最为常见, Mmc 的感染使病羊产生全身多系统综合征,包括胸膜肺炎、关节炎、乳腺炎、结膜炎等^[7-8]。羊群一旦感染,感染率和病死率都很高^[9],对我国的养羊业造成了严重危害。

药物治疗和疫苗接种是防控该病的主要方法。由于药物治疗的效果不尽如人意,传统疫苗本身具有难以克服的局限,所以人们开始致力于新型药物和疫苗的研制。但是目前几乎没有保护性抗原基因的筛选及其详细的研究报道。过去人们大多从表型分析入手,寻找已知功能的编码基因,实际只了解到极少数的基因,极大地阻碍了对支原体的深入研究。近年来,支原体基因组研究取得了飞速的进展。

目前,已经完成 95010、GM12 和 PG3 共 3 株 Mmc 的基因组测序。完整基因组的序列测定带给研究者有关支原体详尽的遗传学和生物学信息,从根本上揭示微生物的全部基因,不仅可发现新的基因,还可发现新的基因间相互作用、新的调控因子等。这一研究将使人类从更高层次上掌握病原微生物的致病机制及其规律,从而得以发展新的诊断、预防及治疗病原体感染的制剂、疫苗及药品。

1 材料与方法

1.1 材料

1.1.1 菌株及来源: 丝状支原体山羊亚种标准株 PG3 (*Mycoplasma mycoides* subsp. *capri* str. PG3, Mmc str. PG3)由内蒙古农业大学兽医学院兽医微生物学与免疫学实验室保藏和提供。

1.1.2 培养基、主要试剂和仪器: 改良的 Hayflick's 培养基,参照郝瑞霞^[10]的方法配制。支原体基因组 DNA 提取试剂盒, Biomiga 公司; Supercoiled DNA Ladder Marker, 宝生物工程(大连)有限公司; 琼脂糖, Bio-Rad 公司。低温高速离心机, Sigma 公司; 核酸电泳仪, 北京百晶生物技术有限公司; UVP 凝胶成像系统, Bio-Rad 公司; 多功能酶标仪, 美国伯腾仪器有限公司。

1.2 方法

1.2.1 菌株复苏和扩大培养: 将真空冷冻干燥保存的供试菌株转接至改良的 Hayflick's 液体培养基中,在 37 °C、5% CO₂ 和一定湿度的环境下培养 5–7 d。以 1:6 的接种比例再连续传代两次,以便菌

株活力得到充分恢复。将复苏后的菌株以 1:10 的接种比例扩大培养。

1.2.2 基因组 DNA 的提取: 取扩大培养的菌液, 参照支原体基因组 DNA 提取试剂盒的说明书提取供试菌株的 DNA。分光光度法检测提取的 DNA 样品的浓度与纯度(OD_{260}/OD_{280})。

1.2.3 基因组测序与注释和分析: 将 Mmc str. PG3 基因组 DNA 样品委托深圳华大基因科技有限公司进行全基因组序列测定、基因组注释、共线性分析、非编码 RNA (ncRNA) 预测和基因岛预测工作。

基因组测序。(1) 基因组 DNA 随机文库的构建: 将基因组 DNA 样品用超声波打断, 电泳回收主带小于等于 800 bp 大小的一系列片段, 用 T4 DNA 聚合酶、Klenow DNA 聚合酶和 T4 PNK 将打断形成的粘性末端修复成平末端, 再通过 3 端加碱基“A”, 使得 DNA 片段能与 3 端带有“T”碱基的特殊接头连接, 用电泳法选择需回收的目的片段连接产物, 再使用 PCR 技术扩增两端带有接头的 DNA 片段, 以高拷贝的 pUC18 为载体, 构建质粒文库。随后将已加入接头的 DNA 片段绑定在含有接头的芯片上, 经 PCR 反应, 将不同片段扩增。(2) 上机测序: 采用 Illumina 公司 2010 年推出的 Illumina HiSeq 2000 测序仪进行测序。Illumina HiSeq 2000 测序仪采用稳定的可逆终止法边合成边测序技术, 使用 4 种含有末端阻断基团和不同荧光信号的碱基进行模板互补链的合成, 不仅确保了测序的高精确性和高顺序性, 而且排除了由重复序列和同聚物导致的测序错误。(3) 序列拼接: 运用 SOAPdenovo 1.05 短序列组装软件对处理后的 Reads 数据进行组装。该软件通过 De Bruijn 作图法, 拼接相似度大于 90% 的 Reads 序列。经多次调整后, Mmc str. PG3 基因组序列组装主要参数 K 设置为 63, 得到最优的组装结果。运用 SOAPAligner 2.20 序列比对软件把 Reads 比对到组装结果上, 统计 Reads 的 Mapping 信息, 对组装结果进行补洞、单碱基校对。利用 Paired-end 关系的 Reads 的 Mapping 信息, 对组装结果进行局部组装和优化。

基因组注释和分析。采用 Glimmer 3.0 软件对

ORF 进行预测, 从组装结果中获得基因序列, 用于后续的基因功能注释分析。基因预测选择长度大于 30 个氨基酸的 ORF, 每个 ORF 用 BLASTp 在蛋白质序列数据库中搜索蛋白质同源性。比较序列和目标序列达到 30% 以上的特异性和具有 60% 以上的同源性序列认为显著相关, 对基因名称和功能进行注解。分别采用 RNAmmer 1.2 和 tRNAscan-SE 1.23 软件对 rRNA 和 tRNA 进行预测, 利用 Rfam 数据库对 ncRNA 进行预测。用软件 MUMmer 3.0 和 Mauve 1.0 对 Mmc str. 95010 和我们所测的 Mmc str. PG3 之间进行全基因组序列比对, 分析其共线性关系。比对时, 以 Mmc str. 95010 为参照序列, Mmc str. PG3 为查询序列。将目标基因组与 GI 库进行 BLAT (Standalone BLAT v. 34) 比对, 基于该比对结果中的比对长度、GI 大小、比对上的 GI 起止位置等信息来确定目标基因组中可能的 GI。将组装结果提交到 IS-finder 网站上 (<https://www-is.biotoul.fr/>) 对 IS 进行预测分析(使用默认参数), 并得出预测结果。

假定可变表面脂蛋白基因的鉴定。(1) 同源性和系统发育分析: 将 Mmc str. PG3 的假定可变表面脂蛋白基因序列和其推测的氨基酸序列分别在 GenBank 数据库中进行 BLASTn 和 BLASTp 同源性比对, 采用软件 DNAMAN V4.1 中 Neighbor-Joining 法构建所测基因序列与高同源性序列的系统发育树, 分析它们的亲缘关系。(2) 基因序列及预测氨基酸序列结构特征: 已经有研究证实 *Mycoplasma mycoides* subsp. *mycoides* Small Colony (MmmSC) 的 *vmm* 和 *Mycoplasma capricolum* subsp. *capricolum* (Mcc) 的 *vmc* 基因编码表达大小和相变异的表面脂蛋白。通过与 *vmm* 和 *vmc* 基因序列及预测氨基酸序列的结构特征的比较, 分析 Mmc str. PG3 的假定可变表面脂蛋白基因序列。

2 结果与分析

2.1 基因组 DNA 的提取与检测

取 2 μ L 提取的基因组 DNA 加入 0.7% 琼脂糖凝胶点样孔中, 在 120 V 电压下电泳 30 min, 电泳结果见图 1。经分光光度法检测, Mmc str. PG3 基

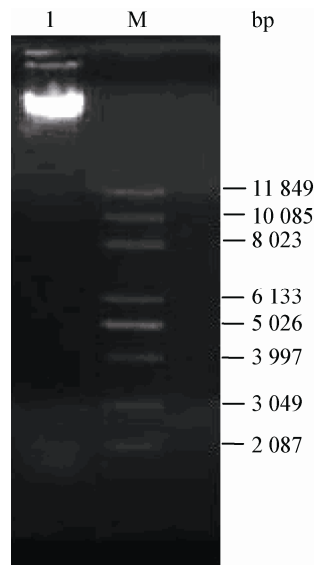


图1 基因组 DNA 凝胶电泳

Figure 1 Genome of Mmc str. PG3

Note: M: Supercoiled DNA Ladder Marker; 1: Genome of Mmc str. PG3.

基因组 DNA 的浓度为 2 440 mg/L, 纯度 OD_{260}/OD_{280} 为 1.81, 达到基因组 DNA 测序要求。

2.2 Mmc str. PG3 基因组注释和分析

2.2.1 Mmc str. PG3 基因组序列基本信息: K-mer 频率分布显示 Mmc str. PG3 预测基因组大小为 1.02 Mb, 错误率为 0.001 2。GC-depth 分析呈泊松分布, G+C% 不存在偏移。SOAPdenovo 1.05 对 Mmc str. PG3 基因组测序所得 Reads 数据进行组装后得到 22 个 Contig, 9 个 Scaffold, N50 和 N90 长度分别为 644 622 bp 和 76 924 bp。基因组大小为 1 025 065 bp, G+C% 为 23.6%, 含有 3 个 rRNA (1 个 5S rRNA, 1 个 16S rRNA, 1 个 23S rRNA), 28 个 tRNA, 12 个 sRNA, 分析预测基因数为

846 个, 编码区大小为 869 436 bp, 基因平均长度为 1 027 bp, 运用 SIGI-HMM 软件在基因组上未检测到 GI。该菌株不存在 IS 或者 ICE 序列。DDBJ/EMBL/GenBank accession ANIV00000000。如表 1 所示, Mmc str. PG3 整个基因组的大小、G+C% 含量和编码基因的数量均小于其他 3 株已完成基因组测序的 Mmc, 只有 Mmc str. 95010 具有质粒, 其他 3 株包括 Mmc str. PG3 没有发现质粒。

2.2.2 Mmc str. PG3 预测基因的 COG 分类: 依据 COG 分类标准将 846 个基因划分为 19 类, 以英文大写字母表示每一类的代码, 即 C-V, 每一类基因的数量和功能见图 2。

2.2.3 Mmc str. PG3 基于 KEGG 代谢通路分类: 通过与 KEGG 数据库进行比对, 将 Mmc str. PG3 的 846 个基因划分为 27 类, 27 类基因序列的表达产物参与 27 个不同代谢通路。每一类基因的数量和功能见图 3。值得注意的是 Mmc str. PG3 基因组中存在由 10 个 ORF 串联而成的麦芽糊精/麦芽糖 (Maltodextrin/Maltose) 基因簇, 参与麦芽糊精/麦芽糖代谢途径, 丝状支原体簇其他成员的基因组中未发现结构与功能相同的基因簇, 见图 4。与 Mmc 的 95010 和 GM12 株相比, PG3 株的 Maltodextrin/Maltose 基因簇缺失了 *TMB* 和 *malC* (MLC1780 和 MLC1790 对应的同源基因), 与 GM12 株一样基因簇上游无插入序列 IS1296。这一证据进一步证明 Maltodextrin/Maltose 基因簇为 Mmc 所特有, 为 Mmc 的检测和由此病原引发的疫病的快速诊断提供了新思路。

2.2.4 基因组共线性分析: 将全基因组序列与已公布的 Mmc str. 95010 的基因组进行比对, 以揭示二

表 1 Mmc str. PG3 基因组与其他 3 株 Mmc 基因组特点比较
Table 1 General features of 3 genomes compared with Mmc str. PG3

菌株 Strain	染色体 Chromosome	质粒 Plasmids	大小 Size (Mb)	G+C% 含量 G+C%	编码基因数量 Number of gene (coding)
95010	1	1	1.16	23.8	924
GM12	1	-	1.09	23.9	911
GM12	1	-	1.08	23.9	908
PG3	1	-	1.02	23.6	846

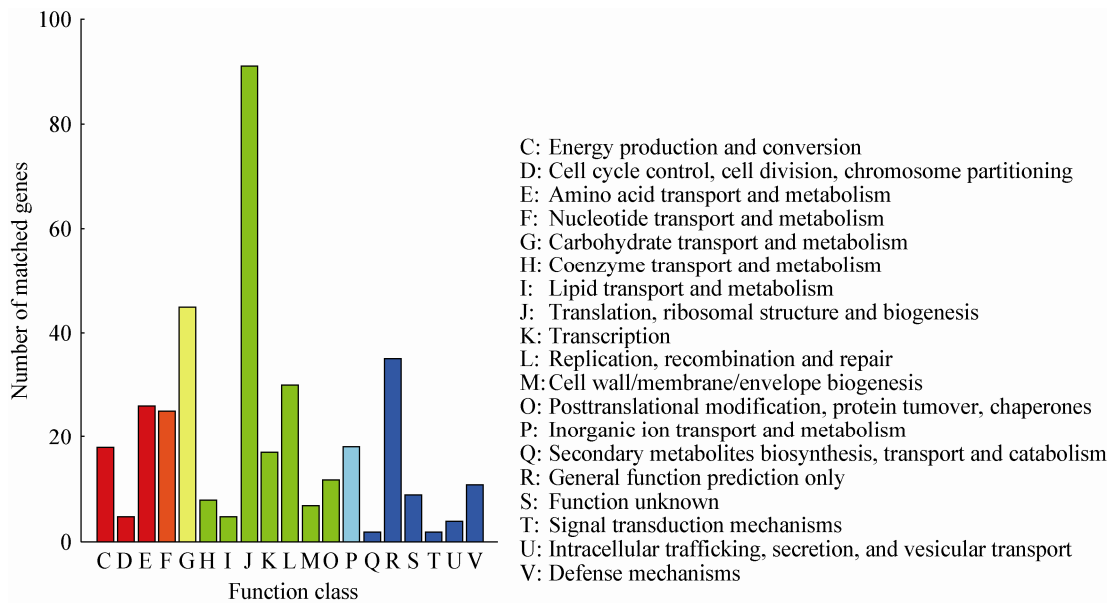


图2 Mmc str. PG3 基因组基因 COG 分类
Figure 2 COG classification for genes of Mmc str. PG3

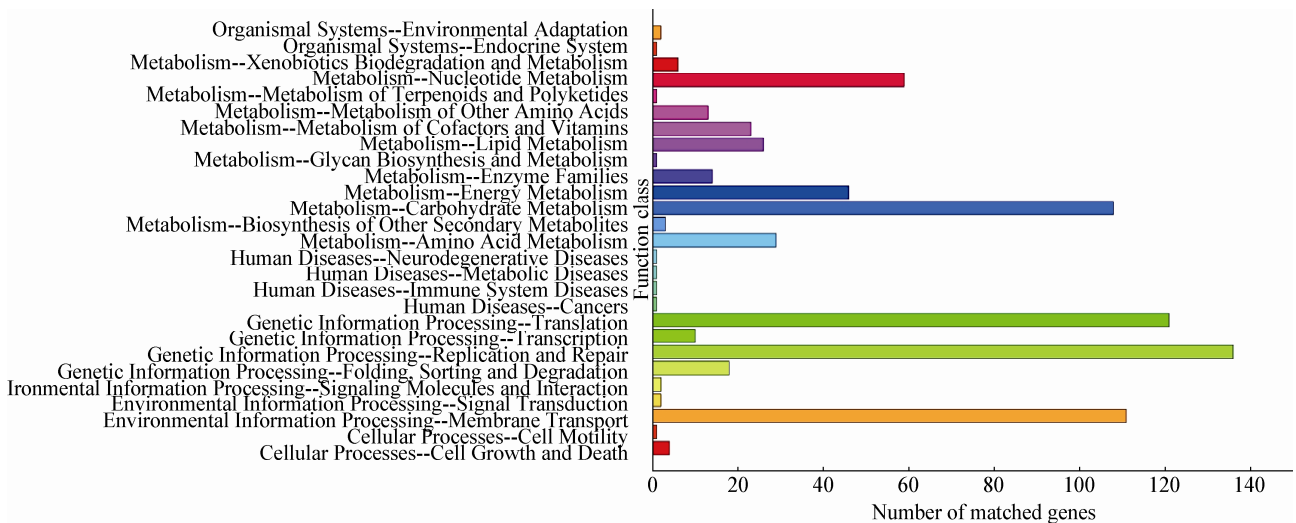


图3 Mmc str. PG3 基因组基因基于 KEGG 代谢通路分类
Figure 3 KEGG pathway classification for genes of Mmc str. PG3

者的亲缘关系远近。比对的结果见图 5 和 6, 从比对结果可以看出二者存在大量的相同序列, 并且保持着良好的线性关系, 没有发生过大规模的插入、缺失、倒位等现象。Mmc str. PG3 和 Mmc str. 95010 很有可能是近期分化的, 亲缘关系较近。这将有助于对 Mmc 引起的 MPSG 的流行病学的研究。

2.2.5 Mmc str. PG3 三个可变表面脂蛋白基因的鉴定: 将 Mmc str. PG3 的 *GL000459*、*GL000461* 和 *GL000462* 三个基因的氨基酸序列在蛋白质序列数据库中搜索, 发现与 Mmc str. GM12 的可变表面脂蛋白 VlcF 同源性最高, 推测具有相似的功能, 因此将这 3 个基因预测为编码可变表面脂蛋白的基

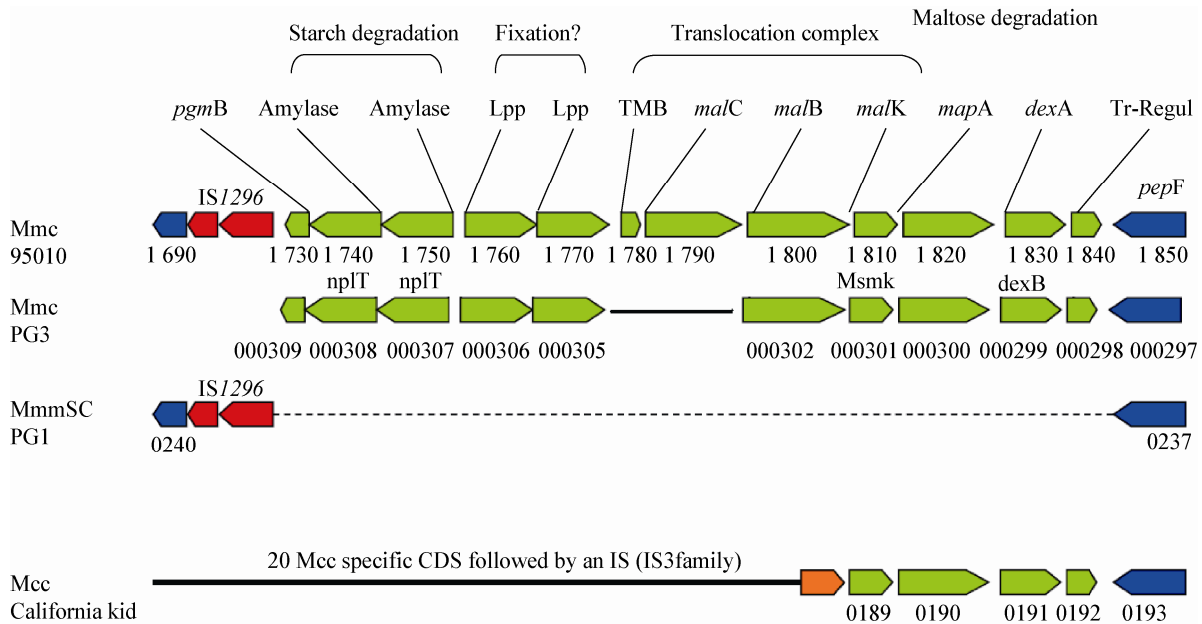


图 4 Mmc str. 95010、PG3、MmmSC str. PG1 和 Mcc str. California kid 的 Maltodextrin/maltose 基因簇
 Figure 4 Organisation of the maltodextrin/maltose gene cluster in the Mmc str. 95010, PG3, MmmSC str. PG1 and Mcc str. California kid

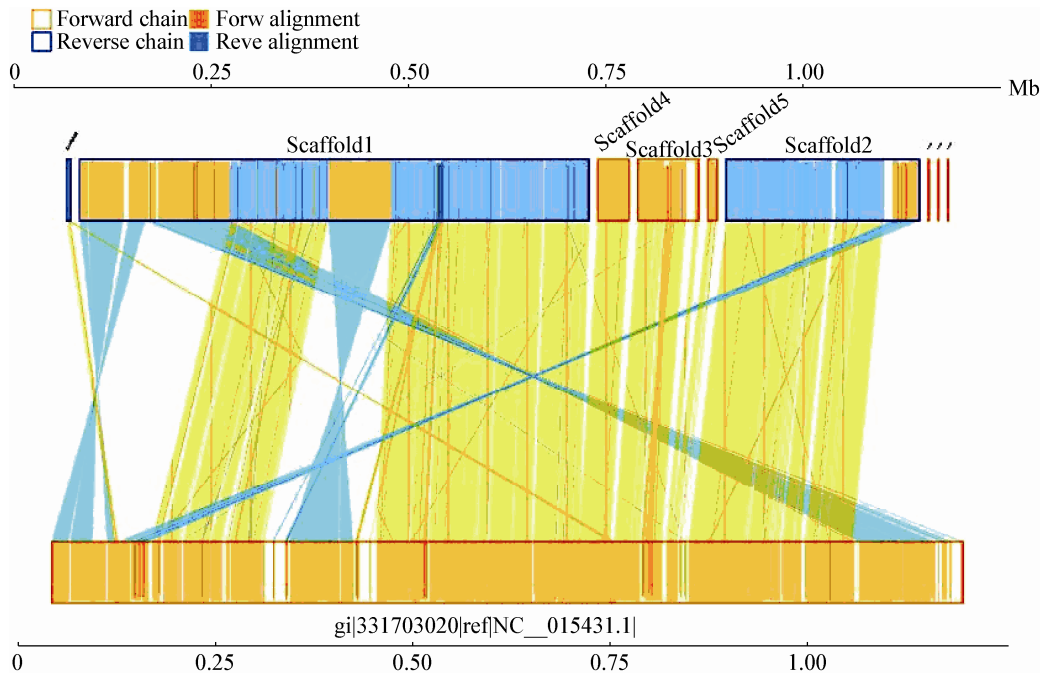


图 5 Mmc str. 95010 和 PG3 的核酸线性共线性图
 Figure 5 Synteny between Mmc str. 95010 and PG3

注：上轴表示参考序列基因组(Mmc str. 95010)，下轴表示所测基因组；上下两条轴中黄色框表示基因组正链，蓝色框表示基因组反链；框中黄色区域表示基因组此区域的核酸序列在正链，蓝色区域表示基因组此区域的核酸序列在负链；图中轴线中间区域连线黄色表示正向比对，蓝色表示反向互补比对。

Note: The upper axis represents the genome of reference sequence (Mmc str. 95010), the lower axis represents the measured genome (Mmc str. PG3); Yellow box in the upper and lower axis indicates forward chain, blue box indicates reverse chain; The yellow region in the box indicates that the nucleic acid sequence of this region is in the positive strand, the blue region indicates that the nucleic acid sequence of this region is in the negative strand; The yellow line in the middle of the axis indicates the forward alignment, the blue line indicates the forward alignment.

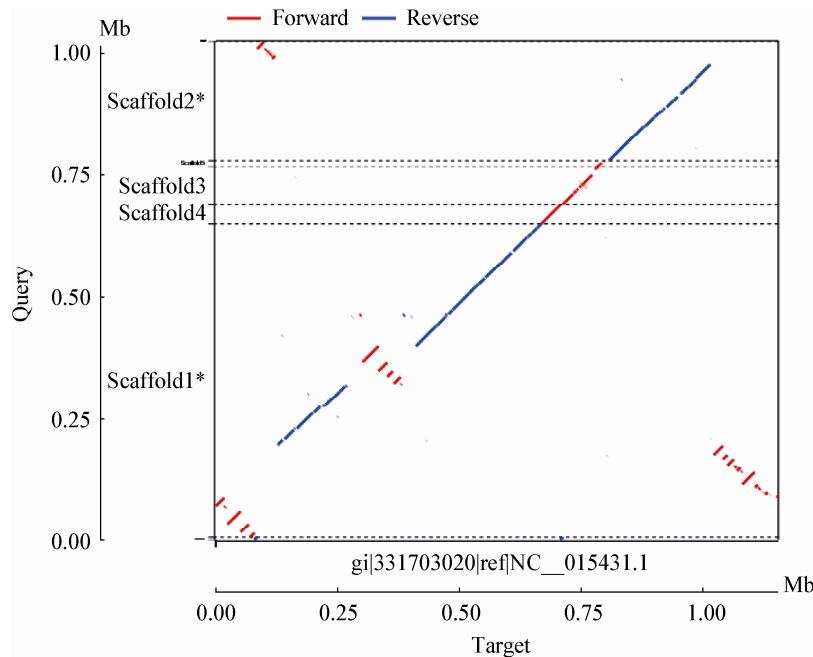


图 6 Mmc str. 95010 和 PG3 全基因组比较散点图

Figure 6 BLAST dot plot of Mmc str. 95010 and PG3

注: 横轴是参考序列基因组, 纵轴是所测基因组; 图中颜色较浅的水平或垂直的直线表示各个 Scaffold 之间的分割(纵轴 Scaffold 右上角*表示所测基因组的负链), 红色线条表示所测基因组的核酸序列在正链, 蓝色表示所测基因组的核酸序列在负链。

Note: The horizontal axis represents the genome of reference sequence, vertical axis represents the measured genome; Lighter color horizontal or vertical straight lines represent division between the various Scaffold (“*” in the upper right corner of Scaffold on the vertical axis represents reverse chain of the measured genome), red line indicates that the nucleic acid sequence of this region is in the positive strand, blue line indicates that the nucleic acid sequence of this region is in the negative strand.

因。(1) 可变表面脂蛋白基因序列的同源性分析: 将可变表面脂蛋白基因序列 *GL000459*、*GL000461* 和 *GL000462* 与 GenBank 数据库的已知序列进行 BLAST 比对, 同源性分析结果显示这 3 个基因序列与 Mm Cluster 中除了 MCCP 外的所有成员(Mmc、Mcc、ML 和 MmmSC)的可变表面脂蛋白基因序列具有亲缘关系。Mmc str. PG3 可变表面脂蛋白基因序列和预测的氨基酸序列与 Mmc str. GM12 的可变脂蛋白基因序列 MMCAP2_0900 和 Mmc str. GM12 的可变脂蛋白氨基酸序列相似性最高, 为 95%和 91%。(2) 可变表面脂蛋白基因序列及预测氨基酸序列结构特征: Mmc str. PG3 可变表面脂蛋白基因 *GL000459*、*GL000461* 和 *GL000462* 在基因组中呈串联连续排列, 基因序列高度同源, 它们的 ORF 分别由 417、624 和 684 个 bp 组成。*GL000459* 基因上游具有典型原核生物的启动子结构, 与 *vmm* 基因的启动子结构高度保守, 由-35 区 TTGACA 盒、

-10 区 TATAAT 盒和 SD 序列(AGGAG)构成, -35 区和-10 区之间的间隔序列为 T(TA)₈, 共 17 个核苷酸;*GL000461* 和 *GL000462* 基因的启动子高度保守, 但没有典型的-10 区 TATAAT 盒和 SD 序列(AGGAG), 在-35 区和-10 区之间均具有 T(TA)_{n-2} 间隔序列, *GL000461* 基因的 TA 重复 10 次; 由于测序原因, 在 *GL000462* 基因启动子 T(TA)_{n-2} 间隔序列与起始密码子之间存在由“N”形成的“洞”, 所以只观察到 3 次 TA 重复, 真实 TA 重复数未知, 见图 7。

Mmc str. PG3 可变表面脂蛋白基因 *GL000459*、*GL000461* 和 *GL000462* 的前脂蛋白分别由 138、207 和 223 个氨基酸组成, 其氨基酸序列结构与 Mm Cluster 中已报道的其他可变表面脂蛋白的氨基酸序列结构高度相似, 信号肽与跨膜区和羧基端的部分序列高度保守, 而羧基端的可变区通常具有重复序列。这 3 个前脂蛋白的氨基酸序列结构特征见图 8, *GL000459*、*GL000461* 的多肽链第 20 个到第

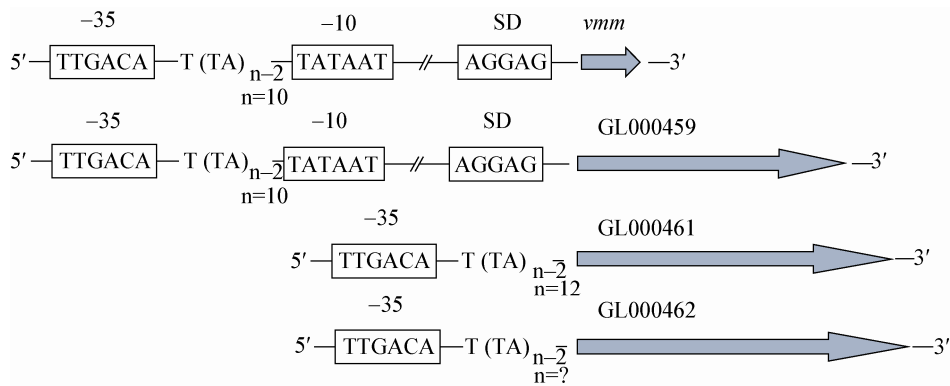


图 7 Mmc str. PG3 可变表面脂蛋白基因结构

Figure 7 Schematic illustrations of three variable surface lipoprotein genes of Mmc str. PG3 with the promoter region and the precursors

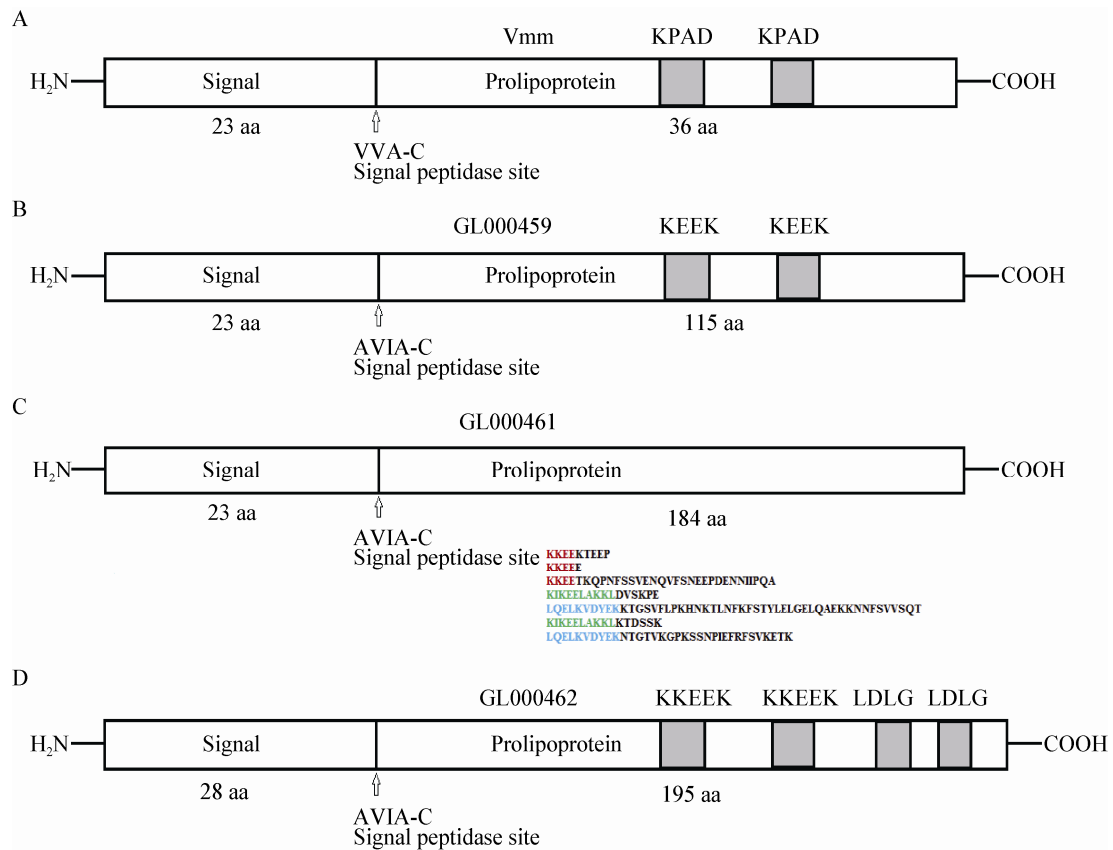


图 8 Mmc str. PG3 可变表面脂蛋白氨基酸序列结构特征

Figure 8 Features of three variable surface lipoprotein precursors of Mmc str. PG3

注: VVA-C、AVIA-C 为前脂蛋白信号肽酶识别位点, 阴影框为由氨基酸组成的重复序列。A 表示 MmmSC str. PG1 的 Vmm 的前脂蛋白氨基酸序列结构特征; B、C、D 分别表示 Mmc str. PG3 的 GL000459、GL000461、GL000462 的前脂蛋白氨基酸序列结构特征。

Note: VVA-C and AVIA-C represent prolipoprotein signal peptidase site; The shaded boxes represent repetitive motifs of amino acid. A represents features of Vmm protein precursor of MmmSC str. PG1; B, C and D represent features of GL000459, GL000461 and GL000462 protein precursors of Mmc str. PG3.

24 个氨基酸, GL000462 的多肽链第 24 个到第 28 个氨基酸, 即 AVIAC, 极可能是前脂蛋白的信号肽酶 II 的识别序列。前脂蛋白若经酶切裂解后, 形成一个信号肽和一个成熟肽序列。

3 个多肽链的成熟肽序列都存在短重复序列。BLASTp 比对结果显示, Mmc str. PG3 前脂蛋白的信号肽与 Mmc str. 95010 和 GM12 可变表面脂蛋白家族成员的信号肽的序列相似性为 91%–96%, 与 MmmSC str. PG1 的 Vmm 的信号肽的序列相似性为 83%–89%, 与 Mcc str. California kid 的 Vmc 可变表面脂蛋白家族成员的信号肽的序列相似性为 63%–82%, 与 ML str. PG50 和 str. 99/014/6 的 Vlc 可变表面脂蛋白家族成员的序列相似性为 74%–96%。而前脂蛋白的成熟肽除了与 Mmc str. GM12 的 VlcF (MMCAP2_0900) 和 95010 的两个可变表面脂蛋白(MLC_9030、MLC_9070)的成熟肽区域序列相似性较高外(77%–91%), 与其他的可变表面脂蛋白序列相似性均低于 50%。由此可见, 这个预测的前脂蛋白的信号肽序列保守, 成熟肽序列变异较大。

3 结论与讨论

我国重庆、贵州和广西等地区均有 Mmc 引起的山羊传染性胸膜肺炎病例的报道。Mmc 的感染导致患病羊萎靡、渐进性消瘦直至死亡, 给养殖户造成了巨大的经济损失, 严重威胁山羊养殖业发展。目前还不清楚 Mmc 的致病机制, Mmc 引起的山羊传染性胸膜肺炎的防治也亟待解决。但是国内外几乎没有 Mmc str. PG3 的毒力因子的相关研究报道, 为了有效防治羊传染性胸膜肺炎, 本研究选择了 Mmc str. PG3 为目标菌株, 对它们的全基因组进行了序列测定和生物信息学分析, 寻找抗原性基因作为研制生物工程疫苗的靶基因序列。Mmc str. 95010 具有五种类型共 24 个 IS 插入序列, 两个特异的结合整合元件 ICE, 每个 ICE 平均长度为 30 kb。并且发现 Mmc 95010 的多基因家族大多有 IS 插入序列。这些可移动的遗传因子很大程度地影响基因组

的可塑性, 导致了大片段 DNA 的水平转移。在 Mmc str. PG3 基因组上未检测到 GI, 该菌株不存在 IS 或者 ICE 序列。在 Mm Cluster 其他成员包括 Mmc str. GM12 中未发现这两个 ICE。由此可见, 只能通过比较基因组的手段来检测 Mmc str. PG3 是否存在水平基因转移现象。

根据 COG 分类和 KEGG 代谢通路分类, Mmc str. PG3 基因组中绝大多数基因主要与蛋白翻译、核糖体结构与合成、DNA 复制与修复、糖代谢和环境信号传递与转换这几个方面有关。Mmc str. PG3 具有 Mmc 特有的麦芽糊精/麦芽糖代谢途径, 提示 Mmc 可能利用麦芽糊精和麦芽糖, 而 Mm Cluster 中其他亚种未发现编码麦芽糊精/麦芽糖代谢途径相关酶类的麦芽糊精/麦芽糖基因簇, 为 Mmc 的病原检测和培养基的改良提供了新思路。

研究者们认为支原体在感染的不同阶段通过表面可变蛋白加强定殖和适应宿主的组织环境。可变表面脂蛋白对于支原体的抗原性、免疫调节和底物结合等有着重要的意义^[11-12]。在 Mmc str. PG3 基因组中预测到 3 个连续排列的可变表面脂蛋白基因, 按照在基因组中的位置命名为 *GL000459*、*GL000461* 和 *GL000462*。由于测序原因, 在 *GL000459* 和 *GL000461* 之间存在由几百个 bp 的“N”形成的“洞”, 生物信息学分析认为这是基因 *GL000460*。可变表面脂蛋白基因一般在基因组中呈串联连续分布, 在 Mmc str. PG3 可变表面脂蛋白基因簇的相应位置能够找到可变表面脂蛋白基因, 推测 *GL000460* 是 Mmc str. PG3 可变表面脂蛋白基因簇的一个成员。这 3 个可变表面脂蛋白基因序列和预测的氨基酸序列与 Mm Cluster 中可变表面脂蛋白基因家族其他成员具有较高的同源性, 这些蛋白的氨基端的信号肽与跨膜区和羧基端的部分序列高度保守, 而羧基端的可变区具有重复序列。其中的一个可变表面脂蛋白基因(*GL000459*)与之前研究的假定可变表面脂蛋白基因(GenBank 登录号: KC170277)^[13]是同一个基因。在进化的过程中发生了由于重复序列的缺失或插入产生的变异现象, 病

原菌可能是通过这种变异逃避宿主的免疫清除而实现对宿主的致病作用。

参 考 文 献

- [1] Kusiluka LJ, Ojeniyi B, Friis NF, et al. Mycoplasmas isolated from the respiratory tract of cattle and goats in Tanzania[J]. Acta Veterinaria Scandinavica, 2000, 41(3): 299-309
- [2] Carmichael LE, George TD, Sullivan ND, et al. Isolation, propagation, and characterization studies of an ovine mycoplasma responsible for proliferative interstitial pneumonia[J]. Cornell Veterinarian, 1972, 62(4): 654-679
- [3] Martrenchar A, Bouchel D, Zoyem N. Isolation and experimental studies of *Mycoplasma mycoides* subsp. *mycoides* LC and *Mycoplasma ovipneumoniae* in goats in Northern Cameroon[J]. Small Ruminant Research, 1995, 16(2): 179-184
- [4] Ojo MO. Caprine pneumonia in Nigeria. I. Epidemiology and bacterial flora of normal and diseased respiratory tracts[J]. Tropical Animal Health and Production, 1976, 8(2): 85-89
- [5] Villalba EJ, Poveda JB, Fernández A, et al. An outbreak caused by *Mycoplasma mycoides* species in goats in the Canary Islands[J]. The Veterinary Record, 1992, 130(15): 330-331
- [6] Edward DG. The pleuropneumonia group of organisms: a review, together with some new observations[J]. Journal of General Microbiology, 1954, 10(1): 27-64
- [7] Alley MR, Clarke JK. The experimental transmission of ovine chronic non-progressive pneumonia[J]. New Zealand Veterinary Journal, 1979, 27(10): 217-220
- [8] Chu YF. An etiological and epidemiological study on contagious caprine pleuropneumoniae in China and development of a bacterin against it[D]. Beijing: Doctoral Dissertation of China Academy of Agricultural Sciences, 2011 (in Chinese)
- [9] 储岳峰. 我国山羊(接触)传染性胸膜肺炎病原学、流行病学研究及灭活疫苗的研制[D]. 北京: 中国农业科学院博士学位论文, 2011
- [9] Deng GM, Zhao X, Liang GX, et al. Diagnosis and control of contagious caprine pleuropneumoniae-like[J]. Chinese Veterinary Science, 1991, 21(6): 3-6 (in Chinese)
邓光明, 赵焯, 梁桂香, 等. 类山羊传染性胸膜肺炎诊断和防治的研究——病原诊断[J]. 中国兽医科技, 1991, 21(6): 3-6
- [10] Hao RX. Isolation identification of *Mycoplasma ovipneumoniae* and genomic sequence determination and analysis[D]. Hohhot: Master's Thesis of Inner Mongolia Agriculture University, 2012 (in Chinese)
郝瑞霞. 绵羊肺炎支原体的分离鉴定及全基因组序列的测定和分析[D]. 呼和浩特: 内蒙古农业大学硕士学位论文, 2012
- [11] Browning GF, Marena MS, Noormohammadi AH, et al. The central role of lipoproteins in the pathogenesis of mycoplasmoses[J]. Veterinary Microbiology, 2011, 153(1/2): 44-50
- [12] Ni B, Bai FF, Liu MJ, et al. Research progress on mycoplasma lipoprotein and its variation interaction with host[J]. Biotechnology Bulletin, 2013, 1(2): 49-54 (in Chinese)
倪博, 白方方, 刘茂军, 等. 支原体脂蛋白及其变异与宿主互作研究进展[J]. 生物技术通报, 2013, 1(2): 49-54
- [13] Xu CG, Gao MH, Zhao YF, et al. Cloning and prokaryotic expression of putative variable surface lipoprotein gene of *Mycoplasma mycoides* subsp. *capri* str. PG3[J]. Progress in Veterinary Medicine, 2013, 34(5): 36-43 (in Chinese)
徐春光, 高明华, 赵艳芳, 等. 丝状支原体山羊亚种假定可变表面脂蛋白基因的克隆及原核表达[J]. 动物医学进展, 2013, 34(5): 36-43