



## 研究报告

## 一株羊源 A 型多杀性巴氏杆菌的全基因组测序及生物学特性分析

张振兴 陈珍 刘昂 程逸文 陈思 杜丽 满初日嘎 王凤阳 陈巧玲\*

海南大学动物科技学院 海南省热带动物繁育与疫病研究重点实验室 海口市动物基因工程重点实验室  
海南 海口 570228

**摘要:**【背景】多杀性巴氏杆菌(*Pasteurella multocida*, Pm)是一种革兰氏阴性菌,可引起动物和人类的呼吸道疾病和败血症等。本实验室前期分离鉴定一株 A 型 Pm HN02 菌株。【目的】通过对 HN02 菌株的全基因组测序及生物信息学分析,扩充多杀性巴氏杆菌的基因组数据库信息;通过毒力基因鉴定和系统进化树分析,明确该菌株含有的毒力基因和遗传进化关系,为临床预防和诊断提供理论依据。【方法】使用单分子实时测序(Single Molecule Real Time Sequencing, SMRT)技术对 Pm HN02 菌株进行全基因组测序,利用 Illumina 测序校正后进行基因功能注释和生物信息学分析。使用 PCR 鉴定菌株毒力基因,并构建进化树进行分析。【结果】Pm HN02 菌株全基因组大小为 2 333 292 bp, GC 含量为 40.15%,预测到的编码基因有 2 389 个,包含 19 个 rRNA (6 个 23S rRNA、6 个 16S rRNA、7 个 5S rRNA)、62 个 tRNA 基因、5 个 sRNA;含 84 个串联重复序列、66 个小卫星 DNA、2 个微卫星 DNA、9 个基因岛、9 个前噬菌体;分别有 1 648、2 190 和 1 917 个基因注释在 GO、KEGG 和 COG 数据库中,而且大部分富集于 Pm 的代谢过程;还有 85 个 III 型分泌系统效应蛋白、191 个表型突变基因、165 个毒力因子相关基因。根据分析结果绘制该菌株的全基因组圈图,并将基因组信息提交至 NCBI 后获得登录号 cp037865。PCR 鉴定发现该菌株含有 *fimA*、*tox A* 等 14 个毒力基因,缺失了 *tadD* 等毒力基因。系统进化树分析发现该菌株同北京的 Pm3 菌株(MH150895.1)进化关系最接近。【结论】研究完成了 A 型 Pm HN02 株的全基因组测序和生物学特性鉴定,揭示了其同国内外 Pm 分离株的进化关系,为预防 Pm 疾病流行和探索 Pm 致病机制提供了参考。

关键词: A 型多杀性巴氏杆菌, 基因组测序, 生物信息学分析, 毒力相关基因, 系统进化树分析

**Foundation items:** Scientific Research Project of Hainan Academician Innovation Platform (YSPTZX202013); China Agriculture Reach System of Mutton Sheep (CARS-38); Central Government of China Guides Local Science and Technology Development Projects (ZY2019HN0903); Research Projects of Universities in Hainan Province (Hnky2021-1)

\*Corresponding author: E-mail: chenqiaoling0520@163.com

Received: 10-02-2021; Accepted: 30-06-2021; Published online: 21-08-2021

基金项目: 海南省院士创新平台科研专项(YSPTZX202013); 国家肉羊产业技术体系(CARS-38); 中央政府引导地方科学与技术发展项目(ZY2019HN0903); 海南省高校科研项目(Hnky2021-1)

\*通信作者: E-mail: chenqiaoling0520@163.com

收稿日期: 2021-02-10; 接受日期: 2021-06-30; 网络首发日期: 2021-08-21

## Whole-genome sequencing and biological characteristics analysis of a sheep-derived capsular type A *Pasteurella multocida* strain

ZHANG Zhenxing CHEN Zhen LIU Ang CHENG Yiwen CHEN Si DU Li  
Manchuriga WANG Fengyang CHEN Qiaoling\*

Haikou Animal Genetic Engineering Key Laboratory, Key Laboratory of Tropical Animal Reproduction & Breeding and Epidemic Disease Research of Hainan Province, College of Animal Science and Technology, Hainan University, Haikou, Hainan 570228, China

**Abstract:** [Background] The Gram-negative *Pasteurella multocida* (Pm) causes respiratory diseases and septicemia in animal and human. Capsular type A Pm HN02 has been isolated and identified by our laboratory. [Objective] We sequenced the whole genome of HN02 and analyzed the bioinformatics, thereby supplementing the genome information of Pm. Through identification and phylogenetic analysis, the virulence genes and genetic evolution of the strain were clarified, respectively, which laid a theoretical basis for clinical prevention and diagnosis of related diseases. [Methods] The whole genome of HN02 was sequenced by SMRT and corrected by Illumina sequencing, followed by genome annotation and bioinformatics analysis. The virulence genes were identified by PCR and the phylogenetic tree was constructed for analysis. [Results] The whole genome of HN02 was 2 333 292 bp, with GC content of 40.15%. It was predicted to have 2 389 coding genes, including 19 rRNA genes (6 23S rRNA, 6 16S rRNA, and 7 5S rRNA genes), 62 tRNA genes, and 5 sRNA genes, 84 tandem repeats, 66 minisatellite DNAs, 2 microsatellite DNAs, 9 genomic islands, and 9 prophages. A total of 1 648, 2 190, and 1 917 genes were annotated in Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Cluster of Orthologous Groups of proteins (COG), respectively. Most of them were involved in the metabolic process of Pm. In addition, 85 type III secretion system effector genes, 191 phenotypic mutation genes, 165 virulence factor-related genes, and 1 important secondary metabolism gene cluster were identified. According to the analysis result, a circular map for the genome was plotted, and the genome information was submitted to NCBI to obtain the accession number cp037865. PCR identification showed that the strain had 14 virulence genes such as *fimA* and *toxA* and deleted virulence genes such as *tadD*. Phylogenetic analysis suggested that HN02 had the closest genetic relationship with the strain (MH150895.1) from Beijing. [Conclusion] The whole-genome sequence and biological characteristics of capsular type A Pm HN02 were elucidated and the evolutionary relationship was revealed, providing a reference for preventing the epidemic of Pm-induced disease and exploring the underlying pathogenic mechanism.

**Keywords:** capsular type A *Pasteurella multocida*, genome sequencing, bioinformatics analysis, virulence-associated genes, phylogenetic analysis

多杀性巴氏杆菌(*Pasteurella multocida*, Pm)是一种革兰氏阴性菌,主要毒力因子有脂多糖(Lipopolysaccharide, LPS)、荚膜、粘附素、外膜蛋白和孔蛋白、多杀性巴氏杆菌毒素等,可引起人类、家畜和许多野生动物的出血性败血症或传染性肺炎<sup>[1]</sup>。Pm引起的动物巴氏杆菌病大多是急性的,主要有出血性败血症、禽霍乱、萎缩性鼻炎、兔鼻塞、猪肺炎和有蹄类动物的下呼吸道感染等。在幼龄、老龄或免疫功能低下的畜群中,

Pm的机会性感染时常发生,其发病率和死亡率都很高,给养殖业和畜牧业带来了巨大的经济损失。此外,Pm还是人类咬伤及软组织感染的常见病原菌,大约50%的猫狗咬伤可导致伤口的Pm感染<sup>[2]</sup>。研究表明,Pm引起的宿主疾病数量有所增加<sup>[3]</sup>,但对于Pm的致病机理尚不清楚。因此,探讨Pm的致病机理具有重要意义。

遗传操作系统的最新发展和多基因组序列的出现,有助于解释特定病理条件下病原菌与特定

宿主的联系, 阐明其致病机制。自 1995 年第一个完整的细菌基因组序列公布以来, 基因组测序得到了蓬勃发展。从传统的 Sanger 测序法到高通量的第二代测序技术, Illumina 平台广受科研人员欢迎, 可被用于细菌的基因组测序。随着技术的发展, PacBio 第三代测序技术于 2011 年推出, 其增加了基因组的测序读取长度, 从而提高了完整基因组的组装速度和质量<sup>[4]</sup>。

本实验室先前从绵羊肺中分离鉴定了一株 A 型 Pm, 命名为 Pm HN02 株。为了进一步探索其致病机理, 本研究使用 SMRT 技术联合 Illumina 测序平台完成 Pm HN02 菌株全基因组测序和基因组的组装, 通过生物信息学分析软件进行基因功能注释, 并使用 PCR 鉴定了菌株的毒力基因, 构建了系统进化树进行分析。

## 1 材料与方法

### 1.1 菌株

A 型多杀性巴氏杆菌 HN02 株分离于辽宁省朝阳市某羊场内呼吸道感染病死绵羊的肺脏, 由本实验室保存<sup>[5]</sup>。

### 1.2 主要试剂和仪器

胰蛋白胨大豆肉汤(Trypticase Soy Broth, TSB)培养基, 青岛海博生物技术有限公司; 细菌基因组提取试剂盒, 天根生化科技(北京)有限公司; NextEra XT DNA 文库制备试剂盒, HiSeq PE Rapid Cluster Kit V2, 北京诺禾致源科技股份有限公司。电泳仪, 北京六一生物科技有限公司; PacBio RS II 系统和 Illumina HiSeq 1500 测序系统, 北京诺禾致源科技股份有限公司。

### 1.3 细菌的培养及全基因组测序

将 Pm HN02 菌株冻存液以 1:100 的体积比加入 TSB 培养基中, 37 °C、180 r/min 培养 10 h。使用细菌基因组 DNA 提取试剂盒提取基因组 DNA, 经琼脂糖凝胶电泳检测和 Qubit 定量质检通过后, 使用 PacBio RS II 系统和 Illumina HiSeq 1500 测序系统进行全基因组测序。

### 1.3.1 Illumina 测序

提取的 Pm HN02 菌株基因组 DNA 通过质检后, 合成基因组测序文库。使用 NextEra XT DNA 文库制备试剂盒制备 DNA 文库, 并在 Illumina 平台(NextSeq)上进行读长 2×150 bp 的测序。简单地说, 1 ng DNA 由 NextEra XT 转座体在 55 °C 下标记 5 min, 然后进行末端修复、加 A 尾、接头连接和文库扩增。

DNA 文库由 Agilent BioAnalyzer 和 Qubit 2.0 系统验证, 用于质量控制分析。文库被变性和稀释到最佳浓度, 并应用于聚类生成步骤。HiSeq PE Rapid Cluster Kit V2 用于在流动细胞上产生群集。使用 Illumina HiSeq Rapid SBS Kit V2 进行 PE151 双端测序, 平均文库大小为 350 bp。

### 1.3.2 PacBio 测序

质检通过后, 使用 PacBio RS II 系统进行测序。根据制造商的说明, 通过使用协议和试剂的 10 kb 模板库制备流程, 从剪切的基因组 DNA 中生成 10 kb 的 SMRTbell 文库, 随后用 PacBio 平台进行测序。

## 1.4 基因组组装

使用 SMRT Link V5.0.1 软件<sup>[6-7]</sup>(<https://www.pacb.com/support/software-downloads/>)将低质量的测序数据过滤, 并将过滤后的测序数据组装成一个没有间隙的重叠群。

## 1.5 基因组注释和生物信息学分析

基因组组分预测包括编码基因、重复序列、非编码 RNA、基因组岛、前噬菌体。

分别使用 GeneMarkS 软件<sup>[8]</sup> (V4.17) (<http://topaz.gatech.edu/GeneMark/>)、phiSpy 软件<sup>[9]</sup> (V2.3) 和 IslandPath-DIOMB 软件<sup>[10]</sup> (V0.2)对基因组进行编码基因、前噬菌体和基因岛预测; 使用 RepeatMasker (V4.0.5)软件<sup>[11]</sup>、Tandem Repeats Finder (TRF, V4.07b)<sup>[12]</sup>进行散在重复序列预测和串联重复序列的预测; 通过 tRNAscan-SE 软件<sup>[13]</sup> (V1.3.1)、rRNAmmer 软件<sup>[14]</sup> (V1.2)对 tRNA 和 rRNA

进行预测；使用 Rfam Database<sup>[15]</sup>和 Cmsearch 程序(V1.1rc4)<sup>[16]</sup>确定最终的 sRNA。

## 1.6 基因功能分析

### 1.6.1 基因功能注释

分别使用数据库 Gene Ontology (GO)<sup>[17]</sup>、Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>[18]</sup>、Cluster of Orthologous Groups of proteins (COG)<sup>[19]</sup>、Non-Redundant Protein Database (NR)<sup>[20]</sup>、Pfam、Transporter Classification Database (TCDB)和 Swiss-Prot 进行基因功能分析。将预测基因的蛋白序列与上述 7 个功能数据库进行 Diamond 比对( $E$  Value $\leq 1e-5$ )；对于每一条序列的比对结果，选取 Score 最高的比对结果(默认 Identity $\geq 40\%$ ，Coverage $\geq 40\%$ )进行注释。

### 1.6.2 毒力基因预测分析

使用 Pathogen Host Interactions (PHI)数据库<sup>[21]</sup>进行病原菌与宿主互作分析；使用 Diamond 软件将目标物种的氨基酸序列与 Virulence Factors of Pathogenic Bacteria (VFDB)数据库<sup>[22]</sup>进行比对，把 Pm HN02 株基因同其相对应的毒力因子功能注释信息结合，得到注释结果；通过数据库 Antibiotic Resistance Genes Database (ARDB)<sup>[23]</sup>、Comprehensive Antibiotic Research Database (CARD)<sup>[24]</sup>可以注释到耐药性相关基因和抗生素耐受种类等。

## 1.7 基因组可视化分析

使用 Circos 软件<sup>[25]</sup>对 Pm HN02 株基因组进行分析并制作基因组圈图。

## 1.8 毒力基因鉴定

参考彭忠<sup>[26]</sup>设计的 23 种多杀性巴氏杆菌毒力基因引物，并参照其体系及程序，以 1.3 中的 Pm HN02 株基因组为模板，进行 PCR 验证。

## 1.9 系统进化树分析

从 Pm HN02 株全基因组测序结果中找到其 16S rRNA 基因序列，在 NCBI 上进行 BLAST 比对，从中挑选 21 条不同地区的同源性序列，利用 MEGA 7.0.26 软件<sup>[27]</sup>中的 Neighbor-Joining 法<sup>[28]</sup>

构建系统进化树。

## 2 结果与分析

### 2.1 Pm HN02 基因组的一般特征

将从 PacBio 和 Illumina 两者产生的序列数据用于从头组装，经过适配接头的修剪和低质量读数的剔除，获得了 Pm 的有效数据，然后使用 SMART Link 中的 Arrow 软件对组装结果优化，并用 Illumina 测序数据进行校正，得到了 Pm HN02 株的全基因组。最后将基因组序列及其注释信息提交至 NCBI 数据库中，获得登录号 cp037865。

Pm HN02 株的全基因组序列全长为 2 333 292 bp，GC 含量为 40.15%，分析得到所有编码基因 2 389 个，所有编码基因的总长度为 2 053 785 bp，平均长度为 860 bp，编码区总长度占全基因的 88.02%；串联重复序列的拷贝数目 84 个，总长度为 5 990 bp，占基因组总长度的 0.256 7%；小卫星 DNA 66 个，微卫星 DNA 2 个；含有 tRNA 基因 62 个，rRNA 基因 19 个(其中 23S rRNA 基因 6 个、16S rRNA 基因 6 个、5S rRNA 基因 7 个)，sRNA 基因 5 个；基因岛 9 个；前噬菌体 9 个，总长度为 299 612 bp，平均长度为 33 290.29 bp。

### 2.2 基因组功能分析结果

将 Pm HN02 株基因的蛋白序列与表 1 中的功能数据库进行比对分析，对于每一条序列的比对结果选取得分最高的比对结果进行注释。Pm HN02 株的全基因组功能注释结果见表 1。

Pm HN02 株的全基因组共有 2 389 个蛋白质编码基因，其中有 2 308、2 190、1 917、1 714、1 648 个基因分别被 NR、KEGG、COG、Swiss-Prot、GO 数据库注释，上述是注释基因数最多的数据库。在 CARD 数据库中注释的基因数目最少为 18 个，在 ARDB 数据库中未注释到基因。

#### 2.2.1 GO 数据库注释结果

在 GO 功能分析中，生物过程中的细胞过程(Cellular Process)和代谢过程(Metabolic Process)是基因富集程度最高的 2 个途径，分别富集了 913 个

表 1 Pm HN02 株基因组功能分析汇总

Table 1 Summary of genome function analysis of the Pm HN02 strain

数据库类型 Database type	基因数量 Gene number
NR	2 308
Swiss-Prot	1 714
KEGG	2 190
COG	1 917
TCDB	370
GO	1 648
PHI	191
VFDB	165
ARDB	0
CARD	18
Secretory protein	124
T3SS	85
CAZY	74

和 930 个基因; 细胞组成中的细胞(Cell)和细胞部分(Cell Part)是基因富集程度最高的 2 个途径, 均富集了 607 个基因; 分子功能中的结合(Binding)和催化活性(Catalytic Activity)是基因富集程度最高的 2 个途径, 分别富集了 754 个和 839 个基因(图 1)。

### 2.2.2 KEGG 数据库和 COG 数据库注释结果

KEGG 代谢通路分析时, 富集在碳水化合物代谢、膜运输、辅助因子、维生素代谢和氨基酸代谢途径的基因比其他途径多(图 2), 这些代谢途径是维持细菌新陈代谢所必需的。

COG 数据库是一种流行的蛋白质功能注释工具, 被广泛用于编码蛋白质的功能分类, 是新基因组测序常规的描述之一。有 1 917 个蛋白质编码基因被注释到 COG 数据库中 A-X 的 23 个类别中(图 3)。在这 1 917 个基因中, 有 205 个与翻译、核糖体结构和生物发生相关的基因, 200 个与碳水化合物运输和代谢相关的基因, 182 个与氨基酸运输和代谢相关的基因, 163 个与细胞壁/膜/被膜生物发生相关的基因, 121 个转录相关基因。这一结果与 KEGG 代谢通路分析的结果相似。大

量基因参与细菌维持基本生命的新陈代谢过程。

### 2.2.3 NR 数据库注释结果

NR 是一个非冗余的蛋白质数据库, 由 NCBI 创建并维护, 其特点在于内容比较全面, 同时注释结果中会包含物种信息, 可用于物种分类。将 Pm HN02 株的基因序列翻译为氨基酸序列后与 NR 数据库进行比对, 有 2 308 个基因在 NR 数据库得到注释, 其中注释到 Pm 上的基因数量最多, 为 2 163 个。

### 2.2.4 TCDB 数据库注释结果

TCDB 转运蛋白分类数据库是膜转运蛋白包括离子通道的分类系统。TCDB 数据库转移系统以 5 个级别进行分类, 第一级统计结果如图 4 所示。

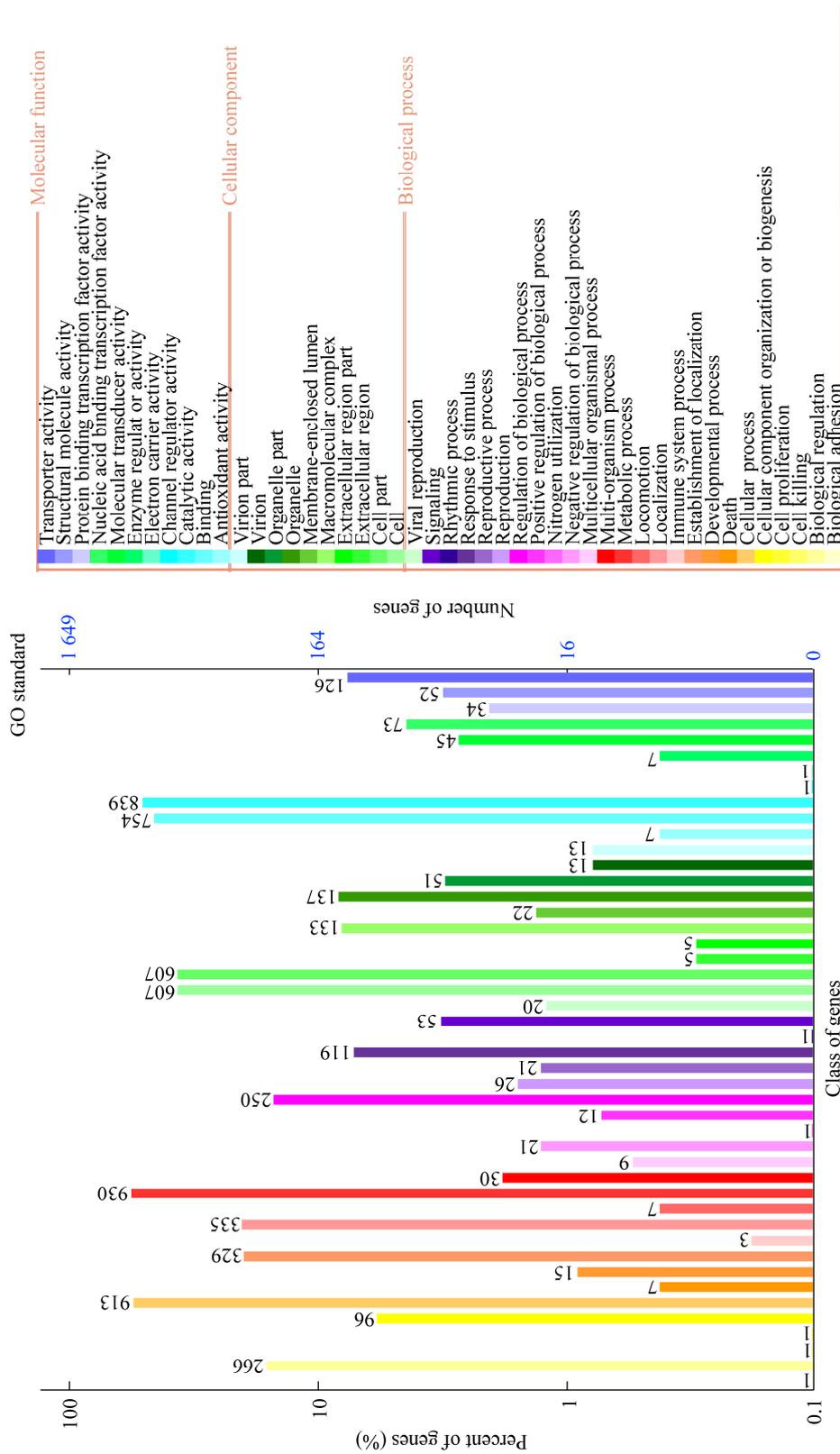
共有 370 个基因注释到 TCDB 数据库的 7 个功能分类中, 它们分别是: Channels/Pores 的 37 个基因, Electrochemical Potential-Driven Transporters 的 99 个基因, Primary Active Transporters 的 161 个基因, Group Translocators 的 20 个基因, Transmembrane Electron Carriers 的 14 个基因, Accessory Factors Involved in Transport 的 6 个基因, Incompletely Characterized Transport Systems 的 33 个基因。

### 2.2.5 Pfam 数据库和 Swiss-Prot 数据库注释结果

蛋白质一般由一个或多个功能区构成, 这些区通常被称为域。结构域的不同组合方式产生的蛋白质在自然界中各不相同。因此, 蛋白结构域的鉴别对分析蛋白质的功能尤其重要。

Pfam 数据库有 2 个组成部分: Pfam-A 和 Pfam-B, 其中 Pfam-A 经过人工筛选, 质量较高。Pm HN02 株中有 1 640 个基因的蛋白序列功能注释到了 Pfam 数据库。

Swiss-Prot 是一个精选的蛋白质序列数据库, 包括蛋白质功能、结构、翻译后修饰、变异等描述, Pm HN02 株中有 1 714 个基因的蛋白序列功能注释到了 Swiss-Prot 数据库。



**图 1 Pm HN02 基因组的 GO 功能分类图**  
**Figure 1 GO function classification map of the Pm HN02 genome**  
 注：横坐标表示样品注释上的 GO 功能分类。右侧纵坐标表示注释上的基因数目；左侧纵坐标表示注释上的基因数目占所有编码基因的百分比  
 Note: The abscissa indicates the GO function classification on the sample annotation. The ordinate on the right indicates the number of genes on the annotation; The ordinate on the left indicates the percentage of the number of genes on the annotation to all coding genes

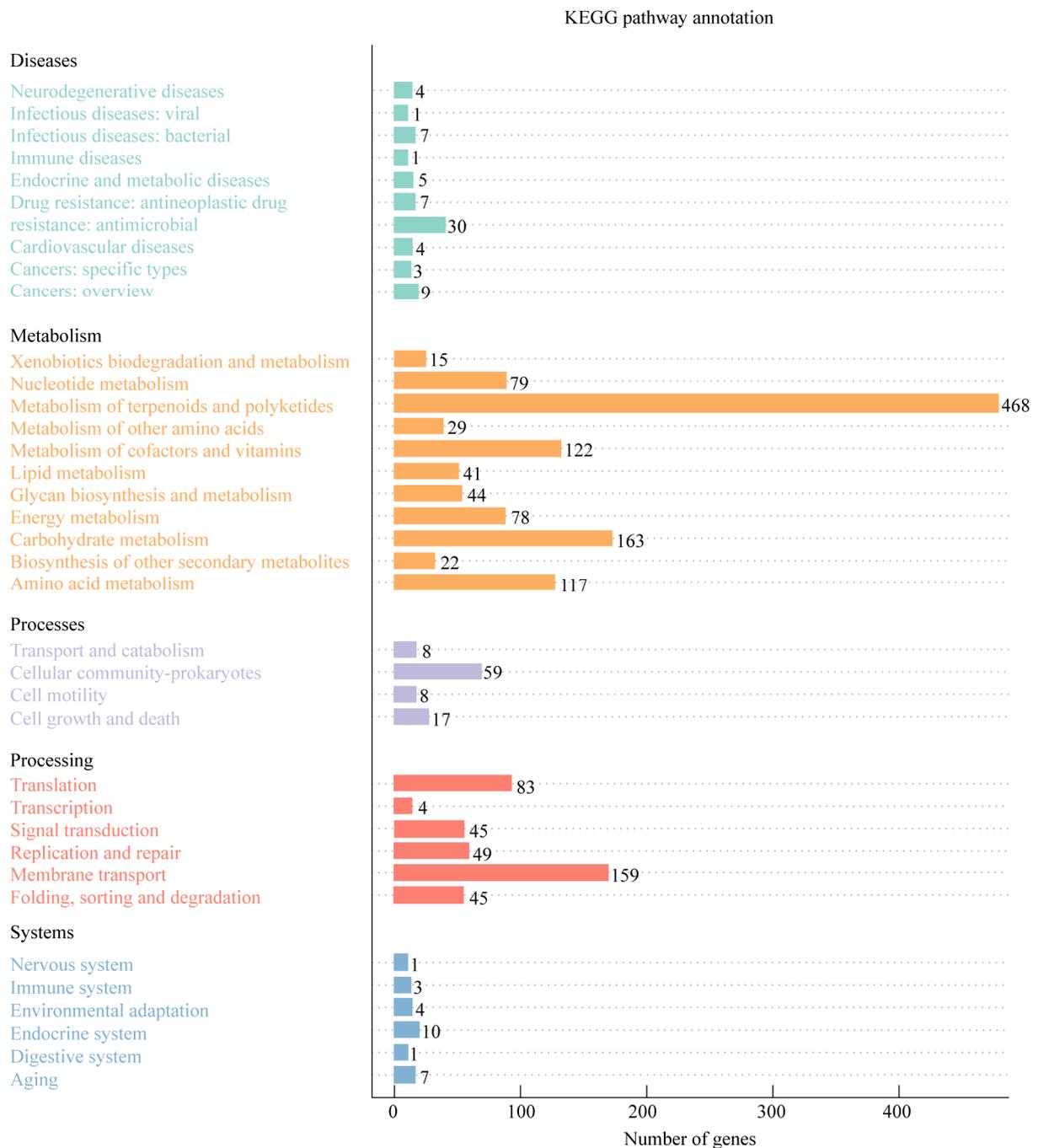


图 2 Pm HN02 基因组的 KEGG 代谢通路分类图

Figure 2 Classification diagram of KEGG metabolic pathway of the Pm HN02 genome

注: 条形图上的数字代表注释到该 KEGG 通路基因数目; 纵坐标轴是数据库中 Level 1 各 KEGG 通路的名称

Note: The number on the bar graph represents the number of genes annotated to the KEGG pathway. The ordinate axis is the name of each KEGG pathway in level 1 in the database

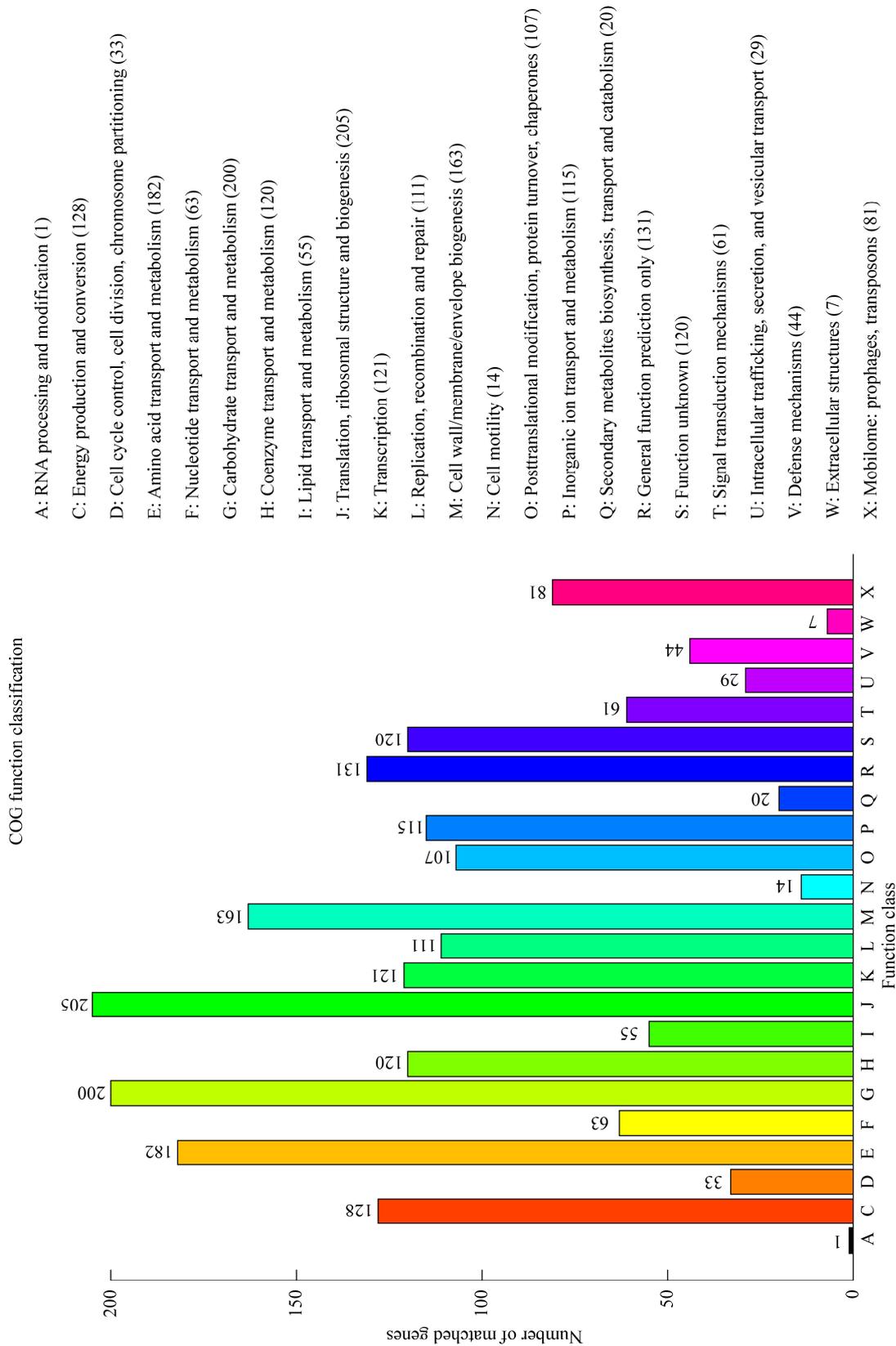


图3 Pm HN02 基因组的 COG 功能分类

Figure 3 COG function classification diagram of the Pm HN02 genome

注: 横坐标表示 COG 功能类型; 纵坐标表示注释上的基因数目

Note: The abscissa indicates the type of COG function. The ordinate indicates the number of genes on the annotation

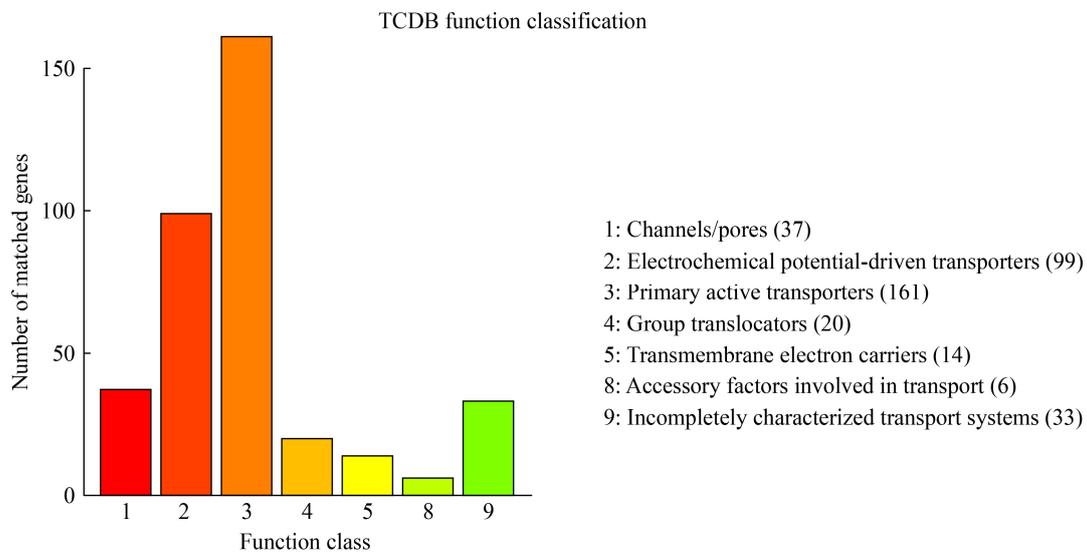


图 4 Pm HN02 基因组的 TCDB 功能分类图

Figure 4 TCDB function classification diagram of the Pm HN02 genome

注: 横坐标表示 TCDB 一级分类类型; 纵坐标表示注释上的基因数目

Note: The abscissa indicates the primary classification type of TCDB. The ordinate indicates the number of genes on the annotation

### 2.2.6 碳水化合物活性酶(CAZy)数据库注释结果

CAZy<sup>[29]</sup>全称为 Carbohydrate-Active Enzymes database, 是碳水化合物酶相关的专业数据库, 内容包括能催化碳水化合物降解、修饰以及生物合成的相关酶系家族; 其包含 5 个主要分类: 糖苷水解酶(Glycoside Hydrolases, GHs)、糖基转移酶(Glycosyl Transferases, GTs)、多糖裂解酶(Polysaccharide Lyases, PLs)、糖类酯解酶(Carbohydrate Esterases, CEs)和氧化还原酶(Auxiliary Activities, AAs); 碳水化合物结合结构域是一种非催化结构域, 能折叠成特定的三维空间结构, 具有结合碳水化合物的功能; 近年来研究表明, 碳水化合物结合结构域能通过结合碳水化合物活性酶的底物, 提高碳水化合物活性酶的催化结构域作用于底物的活性。

Pm HN02 株中有 74 个基因注释到 CAZy 数据库中, 其中包含与碳水化合物相关 Modules 的基因有 14 个、糖类酯解酶的基因 3 个、糖苷水解酶的基因 22 个、糖基转移酶的基因 39 个, 均无多糖裂解酶和氧化还原酶(AAs)基因。

### 2.2.7 病原与宿主互作数据库(PHI)注释结果

PHI 数据库中的每个基因都包含核酸和氨基酸序列及感染宿主过程中预测的蛋白功能的详细描述。Pm HN02 株有 191 个基因注释到 PHI 数据库中, 它们均发生了表型突变, 病原体 PHI 表型突变类型基因数目的统计情况如图 5 所示。

这 191 个基因中有 13 个突变后毒力增加, 123 个突变后毒力降低, 3 个突变后致死, 16 个突变后使菌株丧失致病性, 32 个突变后不受影响, 2 个突变成效应子, 2 个突变后致菌株具有化学耐受性。

### 2.2.8 致病菌毒力基因预测分析结果

VFDB 是一个用于研究病原菌、衣原体和支原体的数据库。除了毒力基因的种类信息和基本特征描述外, 还提供了毒力基因功能和致病机理的详细描述。用 Diamond 软件将 Pm HN02 株基因的氨基酸序列与 VFDB 数据库进行比对, 有 165 个基因得到注释。其中注释到毒力因子脂寡糖(Lipooligosaccharide, LOS)和荚膜(Capsule)的基因是最多的。

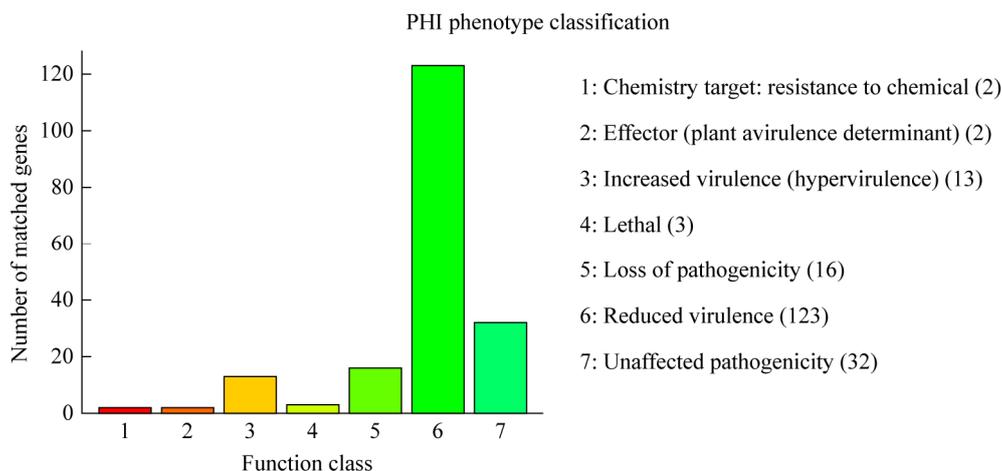


图5 Pm HN02 基因组的 PHI 数据库注释结果

Figure 5 PHI database annotation results of the Pm HN02 genome

注：横坐标表示表型突变类型；纵坐标表示注释上的基因数目

Note: The abscissa indicates the phenotypic mutation type. The ordinate indicates the number of genes on the annotation

### 2.2.9 耐药基因注释结果

使用 ARDB 和 CARD 数据库对 Pm HN02 菌株的所有基因进行注释，发现 Pm HN02 株中只有 18 个基因被注释到 CARD 数据库，但是无基因注释到 ARDB 数据库。

### 2.3 Pm HN02 的全基因组图谱

从 PacBio 和 Illumina 这 2 个平台产生的序列数据被用于从头组装，以生成 Pm HN02 参考基因组。根据基因组序列基本信息、基因预测结果、非编码 RNA 预测结果(sRNA、rRNA、tRNA 注释结果)和生物信息学分析结果(GO、KEGG、COG 注释结果)，利用 Circos 软件绘制该菌株的全基因组图谱(图 6)。

### 2.4 Pm HN02 的毒力基因鉴定结果

Pm HN02 株 23 种毒力基因的 PCR 鉴定结果如图 7 所示。

由图 7 可得，Pm HN02 菌株扩增出了毒力基因 *fimA*、*toxA*、*exxB*、*exxD*、*tonB*、*hgbA*、*tbpA*、*nanB*、*pmHAS*、*ompA*、*oma87*、*plpB*、*sodA* 和 *sodC*，但并未扩增出常见的 *pfhA* 和 *tadD* 等毒力基因。说明在进化过程中，该菌株可能缺

失了一些毒力基因，导致其毒性减弱。

### 2.5 Pm HN02 的遗传进化分析结果

从 Pm HN02 菌株的系统进化树(图 8)可知，Pm HN02 株与中国北京的 Pm3 株(MH150895.1)处在同一分支，进化关系最近；其次同中国新疆的分离株 XJ148-N5-2 株(JQ726521.1)进化关系最近。该菌株分别与中国重庆、湖北、武汉，以及韩国、马来西亚、孟加拉国等地的 Pm 分离株进化关系较近，同苏格兰、丹麦、瑞士、法国、美国等地的 Pm 分离株进化关系较远，符合疾病的地域性流行规律。

## 3 讨论

随着测序技术的快速发展，尽管细菌基因组信息的获取变得越来越容易，然而截止到 2021 年 1 月 29 日，NCBI 数据库中收录了 285 个 Pm 全基因组序列，其中大部分质量偏低，只有少部分是完整的。已收录 Pm 的物种来源很多，但绵羊源很少<sup>[30]</sup>。本研究首次获得了一株羊源 A 型 Pm 的完整基因组信息，扩充了 Pm 的基因组数据库信息。

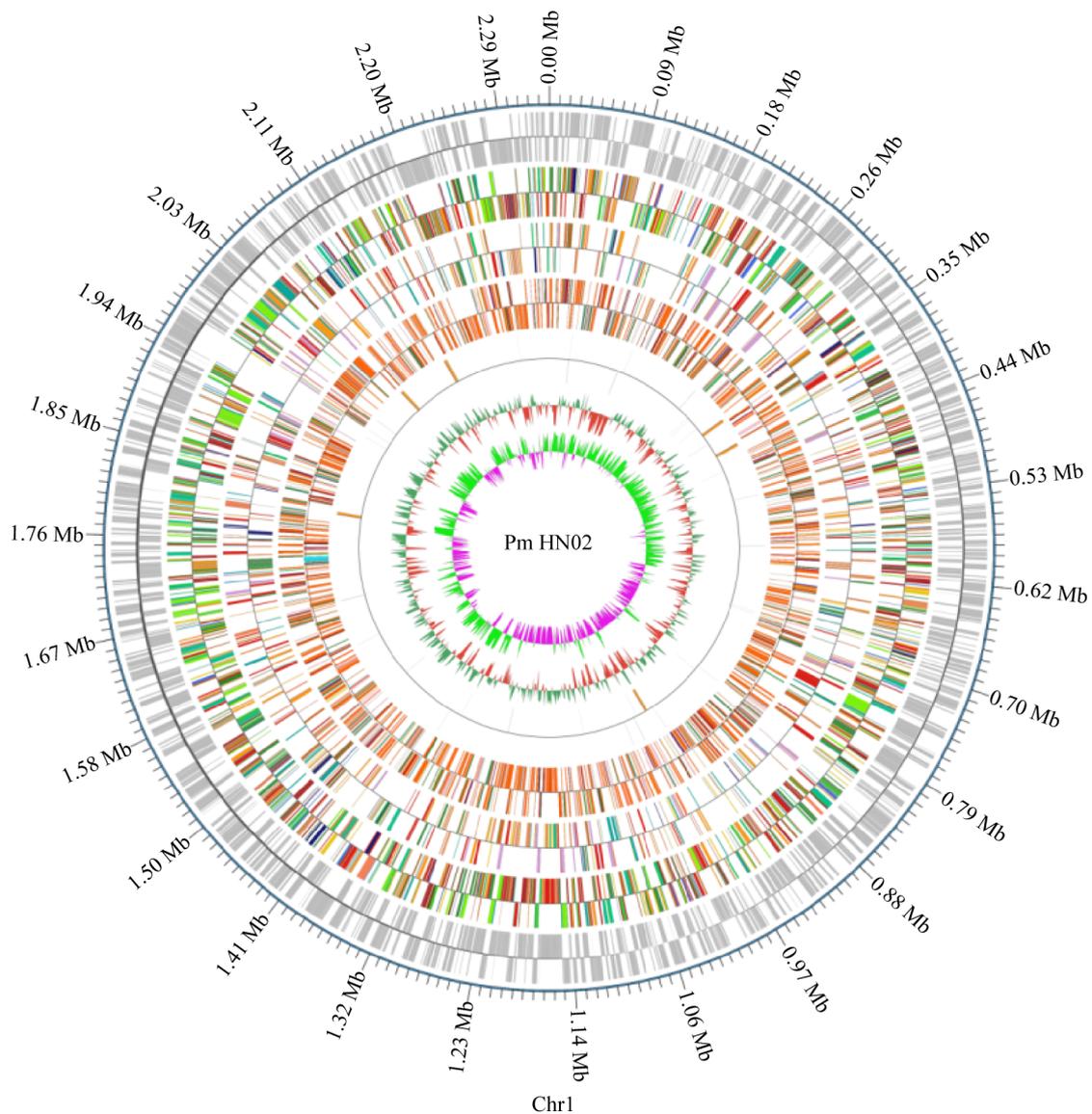


图 6 Pm HN02 的全基因组图谱

**Figure 6 Whole genome map of the Pm HN02**

注: 最外圈是基因组序列位置坐标; 由外到内分别是编码基因、基因功能注释结果(包含 COG、KEGG、GO 数据库的注释结果信息)、ncRNA、基因组 GC 含量、GC Skew 值; 基因组 GC 含量: 以窗口(染色体长度/1 000) bp, 步长(染色体长度/1 000) bp 来统计 GC 含量; 向内的红色部分表示该区域 GC 含量低于全基因组平均 GC 含量, 向外的绿色部分与之相反, 而且峰值越高表示与平均 GC 含量差值越大的基因组; GC Skew 值: 窗口(染色体长度/1 000) bp, 步长(染色体长度/1 000) bp, 具体算法为  $(G-C)/(G+C)$ ; 向内的粉色部分表示该区域 G 的含量低于 C 的含量, 向外的浅绿色部分与之相反

Note: The outermost circle is the position coordinates of the genome sequence. From the outside to the inside, they are coding genes, gene function annotation results (including COG, KEGG, and GO database annotation results), ncRNA, genome GC content, and genome GC skew value. Genome GC content is calculated by window (chromosome length/1 000) bp and step size (chromosome length/1 000) bp. The inward red part indicates that the GC content of the region is lower than the average GC content of the whole genome, and the outward green part. On the contrary, the higher the peak value, the greater the difference from the average GC content. Genome GC skew value: window (chromosome length/1 000) bp, step-size (chromosome length/1 000) bp, the specific algorithm is  $(G-C)/(G+C)$ , the inward pink part indicates that the G content in this region is lower than the C content. The light green part facing outward is the opposite

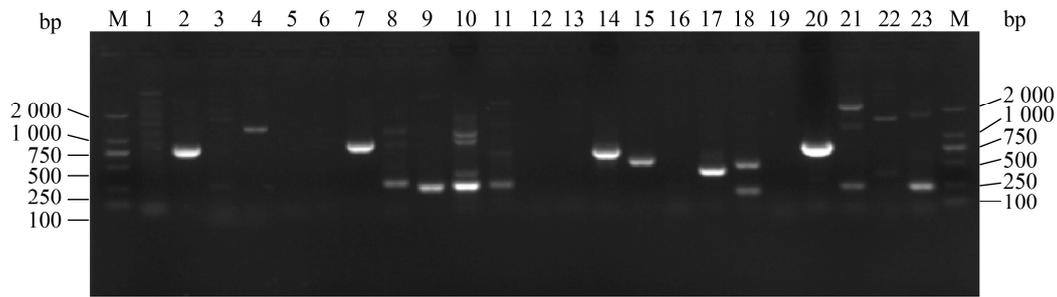


图 7 Pm HN02 株 23 种毒力基因的 PCR 鉴定

**Figure 7 Identification of 23 virulence factor-associated genes of the Pm HN02 strain by PCR**

注: M: DL2000 DNA Marker; 1-23: 毒力基因 *ptfA* (488 bp)、*fimA* (866 bp)、*hsf-1* (654 bp)、*hsf-2* (433 bp)、*pfhA* (286 bp)、*tadD* (416 bp)、*toxA* (864 bp)、*exxB* (283 bp)、*exbD* (247 bp)、*tonB* (261 bp)、*hgbA* (267 bp)、*hgbB* (244 bp)、*Fur* (788 bp)、*tbpA* (728 bp)、*nanB* (555 bp)、*nanH* (287 bp)、*pmHAS* (430 bp)、*ompA* (201 bp)、*ompH* (438 bp)、*oma87* (838 bp)、*plpB* (282 bp)、*sodA* (361 bp) 和 *sodC* (235 bp) 的扩增产物

Notes: M: DL2000 DNA Marker; 1-23: PCR products of virulence genes *ptfA* (488 bp), *fimA* (866 bp), *hsf-1* (654 bp), *hsf-2* (433 bp), *pfhA* (286 bp), *tadD* (416 bp), *toxA* (864 bp), *exxB* (283 bp), *exbD* (247 bp), *tonB* (261 bp), *hgbA* (267 bp), *hgbB* (244 bp), *Fur* (788 bp), *tbpA* (728 bp), *nanB* (555 bp), *nanH* (287 bp), *pmHAS* (430 bp), *ompA* (201 bp), *ompH* (438 bp), *oma87* (838 bp), *plpB* (282 bp), *sodA* (361 bp) and *sodC* (235 bp)

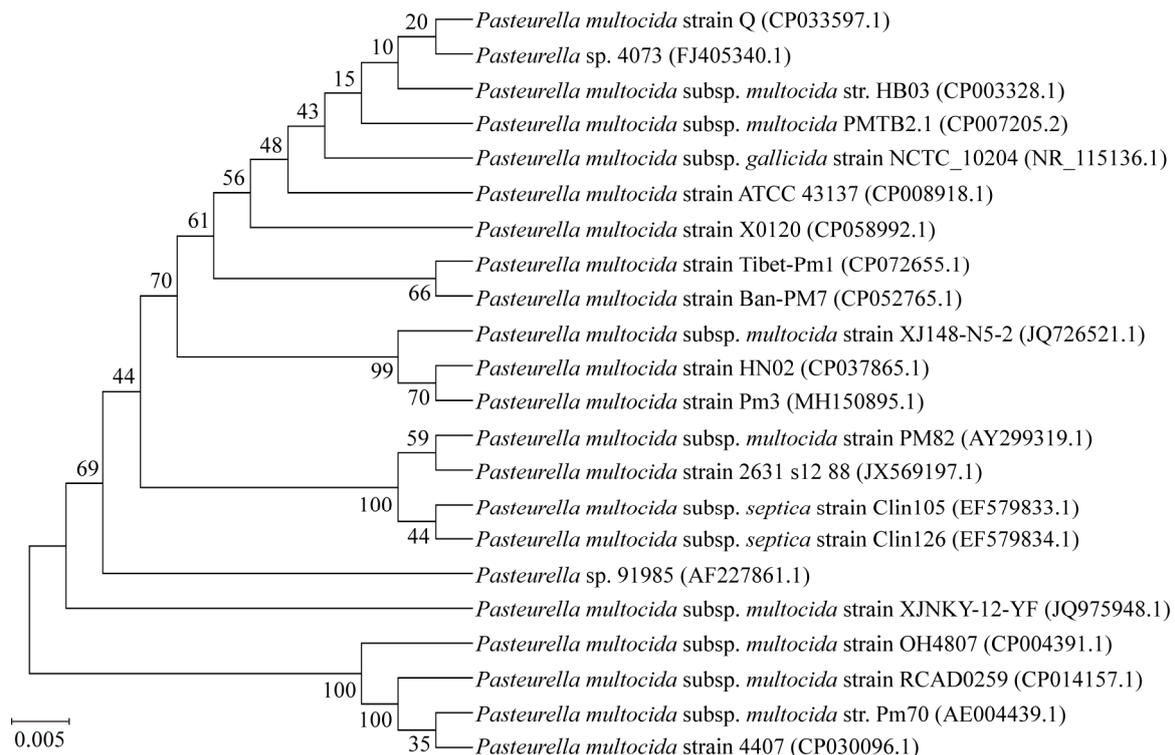


图 8 Pm HN02 菌株基于 16S rRNA 基因序列构建的系统进化树

**Figure 8 Phylogenetic tree of Pm HN02 strain based on 16S rRNA gene sequences**

注: 括号内数字代表各菌株在 GenBank 中的收录号; 比例尺“0.005”代表遗传距离的单位; 每一分支上的数字表示 500 次重复抽样检测的自举分析置信度值

Note: Numbers in parentheses represent the accession number of each strain in GenBank. Scale “0.005” represents the unit of genetic distance. The number on each branch represents the bootstrap analysis confidence value of the 500 repeated sampling test

Pm 是一种常见致病菌, 主要可以引起动物呼吸道疾病, 如慢性上呼吸道感染、急性肺炎或败血症<sup>[2]</sup>等。本研究测序所用的 A 型 Pm HN02 株分离自因呼吸道感染病死绵羊的肺脏。为了进一步探明 Pm 的致病机制, 通过制备 Pm HN02 株的基因组, 利用 SMRT 技术和 Illumina 短读测序相结合的方法进行全基因组测序。其中, SMRT 技术使用 PacBio RS 平台产生更长的读取长度, Illumina 测序确保高精度, 消除极端错误的序列, 并对重复序列和均聚物进行深入测序<sup>[4]</sup>。这 2 种方法相互结合, 高效、准确地完成了对 Pm HN02 全基因组的测序。

Pm HN02 基因组大小为 2 333 292 bp, GC 含量为 40.15%, 与其他 Pm 亚种的基因组相似。Pm HN02 株的全基因组共有 2 389 个蛋白质编码基因, 其中有 1 648、2 190、1 917 个基因分别被 GO、KEGG 和 COG 数据库注释。经过分析发现, COG、GO 和 KEGG 的注释结果相似, 基因富集程度最高的都在代谢过程, 这说明了代谢过程对于维持细菌的生命至关重要<sup>[31]</sup>。只有通过新陈代谢过程, Pm 才能获得足够的营养物质生长和复制。尤其在感染后期, 当 Pm 遇上宿主免疫系统发挥免疫应答, 只有改变参与中枢能量代谢途径基因的表达, 才能更好地吸收铁、氨基酸等营养物质以维持基础代谢存活<sup>[30]</sup>。

用 Diamond 软件将 Pm HN02 株基因的氨基酸序列与 VFDB 数据库进行比对, 有 165 个基因得到注释。其中注释到毒力因子脂寡糖和荚膜 (Capsule) 的基因最多。荚膜由粘性表层组成, 是 Pm 的基本结构成分和致病因子, 在感染过程中为细菌分子模仿、抵抗吞噬和免疫逃避提供了一种免疫学机制, 以保护细菌免受宿主防御<sup>[32]</sup>。脂寡糖是革兰氏阴性菌细胞壁上的一种结构成分, 同脂多糖结构相似, 生物功能也相近, 主要的差别在于脂寡糖没有 O 抗原。研究表明<sup>[33]</sup>, Pm 的脂多糖结构不同于其他革兰氏阴性菌, 其脂多糖缺乏 O 抗原, 被命名为脂寡糖。由于使用 VFDB 进

行毒力因子分析时存在一定的假阳性, 本研究进一步对 Pm HN02 菌株的毒力基因进行了 PCR 鉴定。PCR 结果表明该菌株缺失了 *tadD* 等毒力基因, 说明该菌株在进化过程中缺失了一些毒力基因, 导致其毒性减弱。这为后续研究 Pm 毒力提供了实验基础。

本研究通过构建 Pm HN02 株的 16S rRNA 基因系统进化树, 发现 Pm HN02 株同中国北京的 Pm3 株 (MH150895.1) 处在同一分支, 进化关系最近; 并与中国新疆、重庆、湖北、武汉, 以及韩国、马来西亚、孟加拉国等地的 Pm 分离株进化关系较近; 同苏格兰、丹麦、瑞士、法国、美国等地的 Pm 分离株进化关系较远, 符合疾病的流行规律, 为 Pm 的流行病学调查提供了一定参考。

#### 4 结论

本研究完成了 A 型 Pm HN02 株的全基因组测序和生物学特性鉴定, 揭示了其同国内外 Pm 分离株的进化关系, 为预防 Pm 疾病流行和探索 Pm 致病机制提供了参考。

#### REFERENCES

- [1] Harper M, Boyce JD, Adler B. *Pasteurella multocida* pathogenesis: 125 years after Pasteur[J]. FEMS Microbiology Letters, 2006, 265(1): 1-10
- [2] Wilkie IW, Harper M, Boyce JD, Adler B. *Pasteurella multocida*: diseases and pathogenesis[J]. Current Topics in Microbiology & Immunology, 2012, 361: 1-22
- [3] Harper M, Boyce JD. The myriad properties of *Pasteurella multocida* lipopolysaccharide[J]. Toxins, 2017, 9(8): E254
- [4] Rhoads A, Au KF. PacBio sequencing and its applications[J]. Genomics, Proteomics & Bioinformatics, 2015, 13(5): 278-289
- [5] Cao RY, Zhang ZX, Nie X, Li BB, Huang HF, Yang XJ, Zhu S, Du L, Wang FY. Isolation, identification and phylogenetic analysis of *Pasteurella multocida*[J]. Chinese Journal of Veterinary Medicine, 2018, 54(10): 55-58,3 (in Chinese)  
曹瑞勇, 张振兴, 聂鑫, 李宝宝, 黄海峰, 杨小健, 朱姝, 杜丽, 王凤阳. 多杀性巴氏杆菌的分离鉴定及其系统进化分析[J]. 中国兽医杂志, 2018, 54(10): 55-58,3
- [6] Ardui S, Ameer A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics[J]. Nucleic

- Acids Research, 2018, 46(5): 2159-2168
- [7] Reiner J, Pisani L, Qiao WQ, Singh R, Yang Y, Shi LS, Khan WA, Sebra R, Cohen N, Babu A, et al. Cytogenomic identification and long-read single molecule real-time (SMRT) sequencing of a *Bardet-Biedl Syndrome 9 (BBS9)* deletion[J]. NPJ Genomic Medicine, 2018, 3: 3
- [8] Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions[J]. Nucleic Acids Research, 2001, 29(12): 2607-2618
- [9] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PFAST: a fast phage search tool[J]. Nucleic Acids Research, 2011, 39(suppl): W347-W352
- [10] Hsiao W, Wan I, Jones SJ, Brinkman FSL. IslandPath: aiding detection of genomic islands in prokaryotes[J]. Bioinformatics, 2003, 19(3): 418-420
- [11] Saha S, Bridges S, Magbanua ZV, Peterson DG. Empirical comparison of *ab initio* repeat finding programs[J]. Nucleic Acids Research, 2008, 36(7): 2284-2294
- [12] Benson G. Tandem repeats finder: a program to analyze DNA sequences[J]. Nucleic Acids Research, 1999, 27(2): 573-580
- [13] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence[J]. Nucleic Acids Research, 1997, 25(5): 955-964
- [14] Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes[J]. Nucleic Acids Research, 2007, 35(9): 3100-3108
- [15] Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, et al. Rfam: updates to the RNA families database[J]. Nucleic Acids Research, 2009, 37(Database): D136-D140
- [16] Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments[J]. Bioinformatics, 2009, 25(10): 1335-1337
- [17] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Michael Cherry J, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology[J]. Nature Genetics, 2000, 25(1): 25-29
- [18] Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG[J]. Nucleic Acids Research, 2006, 34(S1): D354-D357
- [19] Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database[J]. Nucleic Acids Research, 2015, 43(D1): D261-D269
- [20] Li WZ, Jaroszewski L, Godzik A. Tolerating some redundancy significantly speeds up clustering of large protein databases[J]. Bioinformatics, 2002, 18(1): 77-82
- [21] Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, De Silva N, Martinez MC, Pedro H, Yates AD, et al. PHI-base: the pathogen-host interactions database[J]. Nucleic Acids Research, 2020, 48(D1): D613-D620
- [22] Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors[J]. Nucleic Acids Research, 2012, 40(D1): D641-D645
- [23] Liu B, Pop M. ARDB: antibiotic resistance genes database[J]. Nucleic Acids Research, 2009, 37(Database): D443-D447
- [24] Jia BF, Raphenya AR, Alcock B, Wagglechner N, Guo PY, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database[J]. Nucleic Acids Research, 2017, 45(D1): D566-D573
- [25] Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics[J]. Genome Research, 2009, 19(9): 1639-1645
- [26] Peng Z. Isolation and whole genome resequencing of *Pasteurella multocida* originated from pigs[D]. Wuhan: Doctoral Dissertation of Huazhong Agricultural University, 2018 (in Chinese)  
彭忠. 猪多杀性巴氏杆菌的分离鉴定及全基因组重测序[D]. 武汉: 华中农业大学博士学位论文, 2018
- [27] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. Molecular Biology and Evolution, 2016, 33(7): 1870-1874
- [28] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees[J]. Molecular Biology and Evolution, 1987, 4(4): 406-425
- [29] Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics[J]. Nucleic Acids Research, 2009, 37(S1): D233-D238
- [30] Peng Z, Wang XR, Zhou R, Chen HC, Wilson BA, Wu B. *Pasteurella multocida*: genotypes and genomics[J]. Microbiology and Molecular Biology Reviews, 2019, 83(4): e00014-19
- [31] Boyce JD, Adler B. How does *Pasteurella multocida* respond to the host environment?[J]. Current Opinion in Microbiology, 2006, 9(1): 117-122
- [32] Petrucci B, Briggs RE, Tatum FM, Swords WE, De Castro C, Molinaro A, Inzana TJ. Capsular polysaccharide interferes with biofilm formation by *Pasteurella multocida* serogroup A[J]. mBio, 2017, 8(6): e01843-e01817
- [33] Hua RQ, Zhao XX, Cheng AC. Research progress in the lipopolysaccharide of *Pasteurella multocida*[J]. Acta Veterinaria et Zootechnica Sinica, 2016, 47(10): 1961-1968 (in Chinese)  
华瑞其, 赵新新, 程安春. 多杀性巴氏杆菌脂多糖的结构与功能研究进展[J]. 畜牧兽医学报, 2016, 47(10): 1961-1968