

项目编号: IPP20108

上海交通大学

第二十期

“上海交通大学大学生创新实践计划” 项目研究论文

论文题目: 临床常见病原菌种类及耐药快速检测研究

项目负责人: 马伟 学院(系): 生命科学技术学院

指导教师: 马伟 学院(系): 生命科学技术学院

参与学生: 曹博 田明振 张商浩

项目执行时间: 2019 年 12 月 至 2020 年 11 月

临床常见病原菌种类及耐药快速检测研究

曹博 致远学院 学号: 518111910092

田明振 致远学院 学号: 518111910021

张商浩 生命科学技术学院 学号: 518111910092

指导老师: 马伟 生命科学技术学院

摘要

细菌是很多人类疾病的病原体,但如何快速确定病原菌的种类一直是一个难题。目前基于传统的细菌分离分析技术,一般需要 5-7 天才能将病原菌种类等重要信息提供给临床医生,这不仅严重影响治疗的疗效,也造成了抗生素的滥用和细菌耐药性的提高。本实验将利用 BLAST 局部比对算法,寻找常见病原菌的特征基因序列,从而建立一个临床常见病原菌种类快速检测体系。我们开发了一套完整、快速且准确的特征基因筛选流程,通过四轮不同层面的筛选,完成了 12 种临床常见病原菌最优特征基因的确定。另外,我们利用大数据组学分析,对引物片段进行筛选,得到了肺炎克雷伯氏菌、大肠埃希菌和铜绿假单胞菌的特征引物对。我们在不同菌株及临床样本中对得到的特征引物进行了特异性验证实验,证明了其很好的特异性和普适性。我们得到的结果对于今后基于基因诊断的病原菌快速检测有很大的帮助,进而可以有效缓解抗生素的滥用和细菌耐药性的提高。

关键词: 特征基因, 特征引物, 病原菌快速检测, BLAST, 生物信息学

ABSTRACT

A Bacteria are the causative agent of many human diseases, but it is always a difficult problem to determine the type of pathogen quickly. At present, based on the traditional bacterial isolation and analysis technology, it generally takes 5-7 days to provide important information such as the type of pathogenic bacteria to clinicians, which not only seriously affects the therapeutic effect, but also leads to the abuse of antibiotics and the increase of bacterial resistance. In this experiment, BLAST local alignment algorithm will be used to find the feature gene sequences of common pathogenic bacteria, to establish a rapid detection system of common pathogenic bacteria in clinical practice. We have developed a complete, rapid and accurate feature gene screening process, and through four rounds of screening at different levels, we have completed the determination of the optimal feature genes of 12 clinically common pathogenic bacteria. In addition, we used big data omics analysis to screen primer fragments, and obtained feature primer pairs of *Klebsiella pneumoniae*, *Escherichia coli* and *Pseudomonas aeruginosa*. We verified the specificity of the feature primers obtained in different strains and clinical samples, which proved their good specificity and universality. The results obtained by us are of great help to the rapid detection of pathogenic bacteria based on gene diagnosis in the future, so as to effectively alleviate the abuse of antibiotics and the increase of bacterial resistance.

Keywords: Feature genes, Feature primers, Rapid detection of pathogenic bacteria, BLAST, Bioinformatics

1. 绪论

1.1 耐药菌感染现状

目前来说,传染性的疾病是世界上最主要的致死原因之一,其中主要是细菌感染以及病毒感染,病毒感染最直接的反应的便是新型冠状病毒 SARS-CoV-2 所引发的新冠疫情的巨大影响,而全世界每年都有超过 1700 万人死于细菌感染。^[1]自 1929 年英国细菌学家弗莱明 (Alexander Fleming) 发现青霉素以来,包括阿奇霉素,头孢菌素等在内的各类抗生素的不断发现和发展改进,各类由细菌引起的感染类的疾病的发病率和死亡率都大幅度下降。然而伴随着人类的不断使用,在基因突变、重组等在内的基因组学的变化,细菌也不断产生着变异,出现了各类的抵抗抗生素的作用机制。目前“ESKAPE”耐药已成为了导致患者发病和死亡的重要原因。^[2]在经济方面,经济学家 Jim O'Neill 主持的 *The Review on Antimicrobial Resistance* 指出,如果任由细菌耐药性的发展,从 2015 年至 2050 年的 35 年中,全球的国内生产总值将会超过 7%,约为 210 万亿美元。而由于耐药性病原体造成的死亡将会达到 1000 万人。^[3]

临床使用的抗生素往往采用的广谱杀菌的方法,即按照实际情况调整相应的用药比例以对大多数细菌造成杀伤作用。而这种广谱杀菌药物的过度使用则进一步加快了细菌耐药基因的产生和全球细菌耐药性发展进程的加快。另一方面,针对细菌耐药性问题的解决方案依旧没有得到有效改善。一方面,由于细菌的突变速率加快和加强,新型有效抗生素的开发更加困难。^[4]另一方面,以噬菌体为代表的新型药物治疗手段,目前仍然处于科研实验阶段,距离临床实验验证和 CFDA/FDA 批准应用相差甚远。

因此在临床上快速实现病原菌的精准鉴定,然后对相应的病菌进行针对性的用药以减缓细菌耐药性的进程是目前能够快速运用起来的关键方式之一。

1.2 病原菌的检测研究现状分析

目前来说病原菌常用的检测方法主要是表型方法 (Phenotypic Methods) 和分子诊断 (Molecular Diagnostic Methods)。其中的表型方法主要是指生化检测 (Biochemical Testing)、显色介质检测 (Chromogenic Media) 和质谱分析 (MALDI-TOF)。对于表型检测的方法一般来说应用于临床方面是利用患者的临床采集得到的样本进行微生物实验室的培养操作,比如划线培养,然后分离单菌落,之后是针对相应的细菌进行检测和鉴定。这种方法应用起来相对较为准确,但是培养周期长,灵敏度和特异性都较低,而周期问题这一点对于临床诊断治疗来说是极为重要的。

分子诊断主要分为 PCR、DNA 微阵列和全基因组测序 (WGS)。分子诊断方法依赖于分析与核酸序列相对应的基因组标记。DNA 微阵列和全基因组测序技术过于复杂,而且费用过高,不利于在基层医院普及。相反,扩增技术操作简便,成本较低,逐渐成为一种优异的检测方法。表 1^[5]展示了目前几种病原菌检测手段的相关属性。目前的技术各有其优点和缺点,它们对细菌检测的适用性取决于它们的特异性、敏感性、经济考虑和仪器的可用性。由此表可以看出 PCR 技术在相关属性中均体现出了优势。

表 1 病原菌检测手段的相关属性 ^{a [5]}

Characteristic	Technique					
	Biochemical testing	Chromogenic media	MALDI-TO F	PCR	DNA microarrays	WGS
Sensitivity	++	++	+++	+++	+++	+++
Specificity	++	+	+++	+++	+++	+++
Cost	++	+	++	+++	++++	++++
Complexity	++	+	++	++	+	++++
Direct detection from clinical samples	NO	YES	NO	YES	YES	NO
Time to result (h)	2-3 ^b	>16	0.2-3 ^b	2	1	>24

^a Within each row, + represents the least and ++++ the most. ^b From previously isolated colonies.

能够应用于病原菌的科学研究方法是极其众多的,但是真正应用到临床检测这一实际问题却是为数不多的。目前仅有三种获得了 FDA 批准(直接测阳性血培养瓶)血流感染的多重检测方法, The Verigene Grampositive blood culture(BC-GP), the FilmArray Blood Culture identification (BCID) Panel, the Verigene Gramnegative blood culture(BC-GN).通过运用一系列的批准方法,以法国生物梅里埃旗下的 Film Array 系统为代表的一代多重 PCR 检测系统现世。这一系统能够从患者的临床症状出发,对引发某一临床症状的最常见的病原体靶标同时进行检测,缩短检测时间,及时控制病情。但是在相应性能的评估中,单一菌株病原体检测准确率为 91.6%,多菌株培养物的检测率仅为 71%。^[6]

理论上来说,该系统能够在血培养瓶报阳之前检测到病原体及其耐药基因,但是该实验系统与步骤检测靶点并没有涵盖所有病原体,检测阳性率尚未达到较高的水平,存在较高的假阳性率。更为重要的是检测成本大幅度提高,难以在普通的中小城市得到推广和普及,而也恰恰是在中小城市中,抗生素耐药性问题出现比例显著高于大城市。因此,通过对临床常见耐药病原菌的快速检测以辅助缓解抗生素耐药性进程的发展这一问题依旧没有得到有效解决。

1.3 研究内容分析

本课题的研究目标定位于呼吸道常见耐药病原菌的检测。我们一共对十二种呼吸道标本分离菌主要菌种肺炎克雷伯氏菌、鲍曼不动杆菌、铜绿假单胞菌、金黄色葡萄球菌、流感嗜血杆菌、肺炎链球菌、嗜麦芽窄食单胞菌、大肠埃希菌、化脓链球菌、阴沟肠杆菌、卡他莫拉菌、军团菌,病原菌的基本情况如表 2 所示。考虑到快速检测试剂盒的研发以及临床鉴定能力,同时预测试剂盒成本及推广程度,这是位于较为适合的一个范围的。在临床鉴定上,通过这些验证这些菌的含量,基本上能够确定绝大多数的病原菌的存在与否。

表 2 十二种呼吸道常见病原菌基本情况信息表

细菌种类 (Bacterial Special)	科属	革兰氏细菌类型	基本形态
肺炎克雷伯氏菌 (<i>Klebsiella pneumoniae</i>)	肠杆菌科 克雷伯氏菌属	革兰氏阴性菌	棒状细菌, 单个或者呈短链, 不运动, 有明显的荚膜
鲍曼不动杆菌 (<i>Acinetobacter baumannii</i>)	莫拉菌科 不动杆菌属	革兰氏阴性菌	圆形杆状, 不具有鞭毛, 移动性不高
铜绿假单胞菌 (<i>Pseudomonas aeruginosa</i>)	假单胞菌科 假单胞菌属	革兰氏阴性菌	长条状, 单个或者成对存在, 无芽孢, 能形成荚膜
金黄色葡萄球菌 (<i>Staphylococcus aureus</i>)	葡萄球菌科 葡萄球菌属	革兰氏阳性菌	球形, 不形成孢子, 不运动
流感嗜血杆菌 (<i>Haemophilus influenzae</i>)	孤菌科 嗜血杆菌属	革兰氏阴性菌	球状杆菌, 有荚膜
肺炎链球菌 (<i>Streptococcus pneumoniae</i>)	链球菌科 链球菌属	革兰氏阳性菌	菌体似矛头状, 双球菌, 有毒株菌体外有荚膜
嗜麦芽窄食单胞菌 (<i>Stenotrophomonas maltophilia</i>)	黄单胞菌科 窄食单胞菌属	革兰氏阴性菌	有 1-8 根极端鞭毛, 有动力、无芽孢、无荚膜
大肠埃希菌 (<i>Escherichia coli</i>)	肠杆菌科 埃希氏菌属	革兰氏阴性菌	短杆菌, 两端呈钝圆形, 具有荚膜结构, 无芽孢
化脓链球菌 (<i>Streptococcus pyogenes</i>)	链球菌科 链球菌属	革兰氏阳性菌	球形或卵圆形, 多数呈链状排列, 无鞭毛, 无芽孢
阴沟肠杆菌 (<i>Enterobacter cloacae</i>)	肠杆菌科 肠杆菌属	革兰氏阴性菌	杆状, 周身有鞭毛, 无芽孢, 无荚膜
卡他莫拉菌 (<i>Moraxella catarrhalis</i>)	莫拉氏菌科 莫拉氏菌属	革兰氏阴性菌	球状, 无芽孢, 无鞭毛



图 1 十二种病原菌的示意图

在本实验过程中，我们建立了临床呼吸道常见病原菌的特征基因的筛选过程。首先利用生物物种进化树来度病原菌进行菌种相关关系进行分析和探讨，随后通过四轮不同的筛选方式，根据我们设计的最初目标逐步完成对病原菌全基因组序列基因的特异性，基因在细菌种属中的覆盖率的筛选工作，建立起此类基因筛查工作的分析体系。在此步骤之后，我们紧接着对相应基因的特征性引物进行了大数据组学分析手段，得到了肺炎克雷伯氏菌、大肠埃希菌和铜绿假单胞菌的特征引物对。在结合特征基因的分析手段的基础之上，我们分别利用单菌株、混合菌株和临床样本对特征引物进行了特异性的验证实验，并筛选得到了最优化的结果。该类结果能够有效应用于绝大多数基于基因诊断的临床呼吸道常见病原菌的快速检测，进而有效缓解抗生素滥用和细菌耐药性提高的发展进程。

2. 研究内容及方法

2.1 病原菌及其亲缘关系的确定

2.1.1 病原菌的选择

我们参考 2017 年 CHINET 中国细菌耐药性监测数据，得到了前 12 种临床常见病原菌，并分析了他们的基本特性（表 3），这 12 种在临床病原菌中的比例高达 90.23%，其中大部分是革兰氏阴性菌，感染比例前三位分别是肺炎克雷伯氏菌、鲍曼不动杆菌和铜绿假单胞菌。[7]

表 3 常见临床病原菌耐药及感染情况

细菌种类 (Bacterial Special)	耐药特性	感染比例
肺炎克雷伯氏菌(<i>Klebsiella pneumoniae</i>)	头孢菌素和氨基糖苷类	19.36%
鲍曼不动杆菌(<i>Acinetobacter baumannii</i>)	氨基糖苷类、环丙沙星	17.00%
铜绿假单胞菌(<i>Pseudomonas aeruginosa</i>)	青霉素类	16.04%
金黄色葡萄球菌 (<i>Staphylococcus aureus</i>)	β -内酰胺类，甲氧西林	7.93%
流感嗜血杆菌(<i>Haemophilus influenzae</i>)	氨苄青霉素	7.06%
肺炎链球菌(<i>Streptococcus pneumoniae</i>)	青霉素、红霉素、四环素类	5.25%
嗜麦芽窄食单胞菌(<i>Stenotrophomonas maltophilia</i>)	β -内酰胺类，氨基糖苷类，喹诺酮类,碳青霉烯类	5.24%
大肠埃希菌(<i>Escherichia coli</i>)	头孢呋辛等头孢类药物	4.74%

化脓链球菌(<i>Streptococcus pyogenes</i>)	青霉素类	3.19%
阴沟肠杆菌 (<i>Enterobacter cloacae</i>)	β -内酰胺酶, AMP C	2.42%
卡他莫拉菌(<i>Moraxella catarrhalis</i>)	氨苄西林	2.00%
军团菌(<i>Legionella</i>)	青霉素, 头孢菌素	/

2.1.2 病原菌基因组的获取

本实验所用到的细菌全基因组数据是由 NCBI 数据库中下载得到的, 各个基因组的具体信息如表 4 所示。每种菌的基因组选用其代表性菌株。(后续会与全部菌株进行普适性筛选)

表 4 细菌基因组信息

菌种	代表性菌株	基因数量
肺炎克雷伯菌	<i>Klebsiella pneumoniae subsp. pneumoniae</i> HS11286	5404
鲍曼不动杆菌	<i>Acinetobacter baumannii</i> ASM74664v1	4043
大肠杆菌	<i>Escherichia coli str. K-12 substr. MG1655</i>	4566
肺炎链球菌	<i>Streptococcus pneumoniae</i> R6	1967
化脓链球菌	<i>Streptococcus pyogenes</i> M1 GAS	1801
金黄色葡萄球菌	<i>Staphylococcus aureus subsp. aureus</i> NCTC 8325	2872
嗜肺军团菌	<i>Legionella pneumophila subsp. pneumophila str. Philadelphia 1</i>	3119
卡他莫拉菌	<i>Moraxella catarrhalis</i> BBH18	1738
流感嗜血杆菌	<i>Haemophilus influenzae</i> Rd KW20	1765
嗜麦芽窄食单胞菌	<i>Stenotrophomonas maltophilia</i> K279a	4490
铜绿假单胞菌	<i>Pseudomonas aeruginosa</i> PAO1	5697
阴沟肠杆菌	<i>Enterobacter cloacae</i> ASM298219v1	5023

2.1.3 系统发育树的构建

16S rRNA 为原核生物的一种核糖体 RNA。由于其基因序列的保守性和存在的普遍性, 16S rRNA 可作为微生物进化距离和亲缘关系的分子指标, 被称为“细菌化石”。^[8]根据 16S rRNA 基因序列构建 12 种病原菌的系统发育树, 可以确定各种细菌间的亲缘关系, 从而确定细菌的最近亲缘菌种, 为后续的特异性筛选提供指导。本项目所用到的 16S rRNA 基因由 NCBI 数据库中下载得到的, 系统发育树使用 MEGA 软件制作。

2.2 病原菌特征基因的筛选与验证

为了寻找临床常见病原菌的特征序列, 我们利用 BLAST 开发了一套完整、快速且准确的特征基因筛选流程, 通过四轮不同层面的筛选, 完成了 12 种临床常见病原菌最优特征基因的确定, 其中多种病原菌的特征基因具有非常好的特异性和普适性。四轮筛选分别为: 预筛选、特异性筛选、普适性筛选和最优化筛选。筛选流程如图 2 所示。

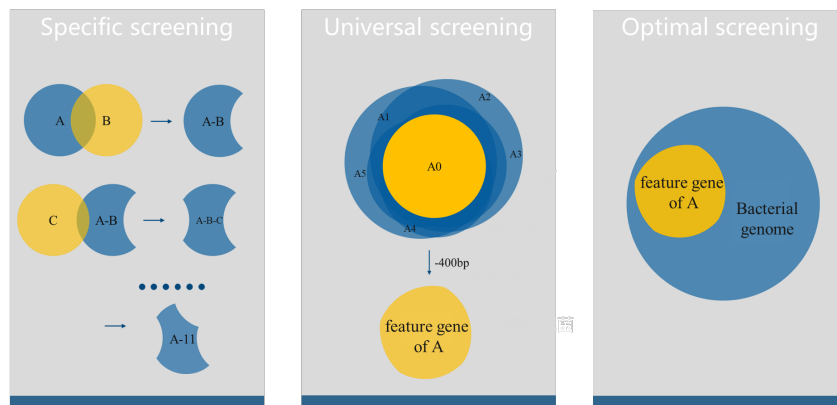


图 2.特征基因筛选流程

2.2.1 预筛选

为了验证筛选方法的可靠性和可行性，我们进行了预筛选，具体实施方法是用一株 A 菌的全基因组与和其亲缘关系最近的 B 菌种的基因数据库进行 BLAST 局部比对，得到二者的匹配序列，将匹配到的序列从原始基因组中剔除，即可得到 A 菌株区别于 B 菌种的特征基因。

2.2.2 特异性筛选

第一轮预筛选剔除了大部分的相似基因，得到了少量的特征基因，证明了此筛选方法的可行性和可靠性。但仅凭第一轮筛选结果还不足以作为临床检测参考，因为这些特征基因只可以区别其亲缘菌种，但不能区别于其他临床常见病原菌，可能造成多个阳性结果。所以我们进行了第二轮筛选-特异性筛选。

我们将 A 菌株预筛选获得的特征序列与其余 10 种细菌的基因组两两进行 BLAST 局部比对，将匹配到的序列从特征序列中多次剔除，最终得到了 A 菌株区别于其他 11 种细菌的特征基因。

2.2.3 普适性筛选

经过前两轮筛选，我们基本上确定了一株菌相对于其他病原菌的特征基因，但由于我们所用的是病原菌的代表性菌株，不能确定已经获得的特征基因是否在所有菌株中都存在。为了提高我们获得的特征基因的普适性，我们进行了第三轮筛选-普适性筛选。

我们将 A 菌株在前两轮中获得的特征基因与这一菌种的所有菌株进行 BLAST 局部对比，剔除没有匹配序列的基因，即只在 A 菌株中存在而其他同种株不存在的基因，从而使特征基因具有最基本的普适性。此轮筛选还剔除了序列长度小于 400bp 的基因，因为序列长度过小会影响引物的设计。

2.2.4 最优化筛选

经过第三轮筛选，我们已经得到了特异性和普适性都比较好的特征基因。为了从这些基因种挑选出最能用于临床实验的特征基因，我们进行了第四轮筛选-最优筛选。

我们将第三轮获得的基因与细菌全基因组进行 BLAST 局部比对，通过分析比对结果，挑选了 3-5 个特异性最强、普适性最好的特征基因。特异性由结果中杂菌数量最少决定，普适性由结果中同种菌株数量最多(organisms)决定。

2.2.5 功能初步研究

我们将获得的最优特征基因在 NCBI 数据库中进行搜索，初步得到了基因所编码蛋白质的简单功能，并对功能与特异性的关联与合理性进行了分析。

2.2.6 引物与探针的设计和筛选

我们根据得到的特征基因设计了多个引物和探针，并进行了扩增效率测试，筛选出了扩增能力较好的引物和探针。

2.3 病原菌特征引物的筛选与验证

2.3.1 病原菌特征引物筛选算法

在对设计的引物进行多菌种特异性验证之后，我们发现其效果并未达到预期特异性和准确性，于是我们选择直接对引物片段进行筛选。为此我们利用大数据组学分析开发了以下算法，如图 3 所示。首先我们对待选菌株和背景菌株的基因序列进行格式化和 N 元切分，并根据引物通用规则进行了过滤，分别得到了他们的候选引物序列数据库。接着，我们通过多轮单、双序列全匹配普适度和特异性过滤，得到了具有特异性和普适性的特征引物对。由于病原菌全基因组数据

量庞大，可达百亿量级，常规的暴力筛选方法耗时惊人。为此我们添加了布隆过滤器、模糊算法等功能，优化了筛选路径，极大地提高了运行速度。我们在筛选的过程中，对引物对的特异性和普适性进行了量化评分，并使用 BLAST 方法对得到的引物进行了分析，最终得到了每种菌株的最优特异性引物对。

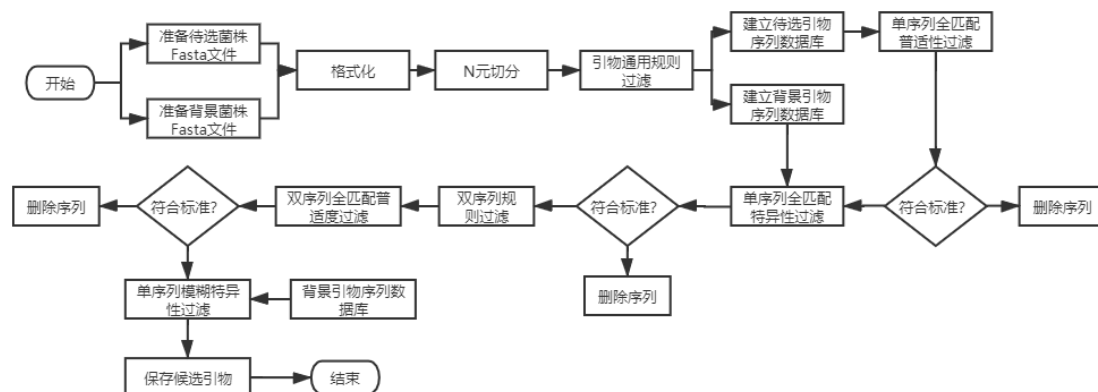


图3 筛选特征引物对算法流程图

2.3.2 实验验证特征引物与探针

本项目致力于利用生物信息学的方法开发出一套高效、准确的临床常见病原菌基因组特征基因与特征引物挖掘机制，利用核酸体外扩增技术实现对临床常见病原菌的快速检测。目前本项目已完成对肺炎克雷伯氏菌、大肠杆菌、铜绿假单胞菌等呼吸道感染临床常见病原菌的特征基因与特征引物挖掘工作。利用核酸扩增技术实现对临床常见病原菌的快速检测的关键在于核酸体外扩增时所用的特征引物是否具备针对目的菌种的种内普适性与种间特异性。为验证采用我们的方法挖掘出的特征引物具备良好的种内普适性与种间特异性，我们首先提取了肺炎克雷伯氏菌、大肠杆菌、铜绿假单胞菌等细菌对应标准菌株的基因组和已经过鉴定的呼吸道感染临床感染样本经平板稀释涂布法分离纯化所得单菌落的基因组并利用我们所设计的引物针对这些基因组进行 PCR 扩增，对扩增后的扩增产物进行琼脂糖凝胶电泳，根据电泳后是否出现对应条带判断引物是否具备良好的种内普适性与种间特异性。

同时为验证采用新型核酸体外扩增技术重组酶介导的等温扩增技术（即 RAA 技术）进行扩增检测的可行性，本项目对肺炎克雷伯氏菌常见耐药性相关基因 KPC-2 设计了 RAA 引物与荧光探针，根据扩增后产物能否通过琼脂糖凝胶电泳得到对应条带以及对 RAA 扩增过程中荧光强度进行检测确定采用 RAA 技术进行扩增的可行性，同时确定采用 RAA 技术为基础的联合荧光探针的目标基因检测技术所能达到的灵敏度。

3. 研究结果

3.1 系统发育树的构建

使用 MEGA 软件对 12 种病原菌的 16 S rRNA 基因进行分析，绘制其系统发育树，如图 4 所示。根据此结果我们得到了 12 种病原菌的亲缘关系，并确定了其最近亲缘菌种，作为预筛选的对比菌种，如表 5 所示。根据此方法确定的筛选路径进行预筛选，可以在第一次就过滤掉大部分的相似基因，减轻了后续的计算量。

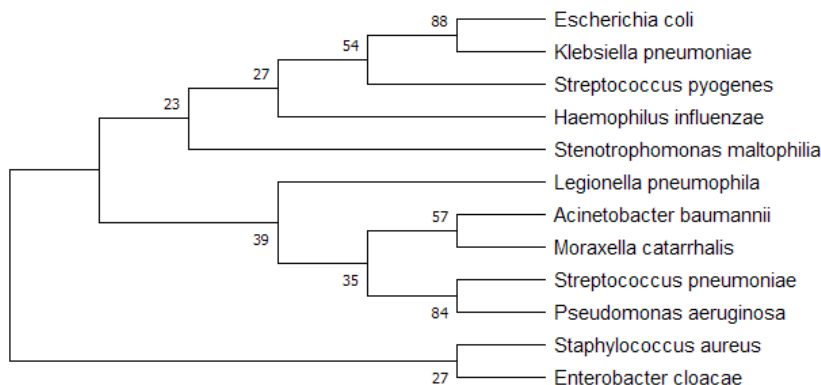


图 4 12 种病原菌的系统发育树

表 5.病原菌亲缘关系

菌种	分类学关系	第一轮筛选对比菌种
肺炎克雷伯氏菌	肠杆菌科，与大肠杆菌、沙门氏菌同科	大肠杆菌 <i>Escherichia coli</i>
鲍曼不动杆菌	假单胞目，和铜绿假单胞菌同目。莫拉菌科，与卡他莫拉菌同科	卡他莫拉菌 <i>Moraxella catarrhalis</i>
大肠杆菌	肠杆菌科，与大肠杆菌、肺炎克雷伯菌同科	肺炎克雷伯菌 <i>Klebsiella pneumoniae</i>
肺炎链球菌	同属链球菌科，链球菌属	铜绿假单胞菌 <i>Pseudomonas aeruginosa</i>
化脓链球菌	同属链球菌科，链球菌属	肺炎克雷伯菌 <i>Klebsiella pneumoniae</i>
金黄色葡萄球菌	芽孢杆菌目，葡萄球菌科，葡萄球菌属	阴沟肠杆菌 <i>Enterobacter cloacae</i>
嗜肺军团菌	军团菌目，军团菌科，军团菌属	鲍曼不动杆菌 <i>Acinetobacter baumannii</i>
卡他莫拉菌	假单胞目，与铜绿假单胞菌同目，莫拉氏菌科，莫拉氏菌属，与鲍曼不动杆菌同科	鲍曼不动杆菌 <i>Acinetobacter baumannii</i>
流感嗜血杆菌	巴氏杆菌目，巴氏杆菌科，嗜血杆菌属	化脓链球菌 <i>Streptococcus pyogenes</i>
嗜麦芽窄食单胞菌	黄单胞菌目	省略邻近菌种对比
铜绿假单胞菌	假单胞菌目，假单胞菌科，假单胞菌属	肺炎链球菌 <i>Streptococcus pneumoniae</i>
阴沟肠杆菌	肠杆菌属，与大肠杆菌同属	金黄色葡萄球菌 <i>Staphylococcus aureus</i>

3. 2 特征基因筛选结果及引物设计

3. 2. 1 特异性筛选结果

将 A 菌株预筛选获得的特征序列与其余 10 种细菌的基因组两两进行 BLAST 局部比对，将匹配到的序列从特征序列中多次剔除，最终得到了 A 菌株区别于其他 11 种细菌的特征基因。结果如表 6 所示。

表 6.特异性筛选结果

菌种	代表性菌株	基因数量	剔除基因数	特征基因数	滤过率
肺炎克雷伯氏菌	<i>Klebsiella pneumoniae subsp. pneumoniae</i> HS11286	5404	5343	61	98.87%
鲍曼不动杆菌	<i>Acinetobacter baumannii</i> ASM74664v1	4043	3790	253	93.74%
大肠杆菌	<i>Escherichia coli str. K-12 substr. MG1655</i>	4566	4369	197	95.69%
肺炎链球菌	<i>Streptococcus pneumoniae</i> R6	1967	1910	57	97.10%
金黄色葡萄球菌	<i>Staphylococcus aureus subsp. aureus</i> NCTC 8325	2872	2829	43	98.50%
嗜肺军团菌	<i>Legionella pneumophila subsp. pneumophila str. Philadelphia 1</i>	3119	2588	531	82.98%
流感嗜血杆菌	<i>Haemophilus influenzae</i> Rd KW20	1765	1664	101	94.28%

铜绿假单胞菌	<i>Pseudomonas aeruginosa</i> PAO1	5697	5688	9	99.84%
卡他莫拉菌	<i>Moraxella catarrhalis</i> BBH18	1738	1629	109	93.73%
嗜麦芽窄食单胞菌	<i>Stenotrophomonas maltophilia</i> K279a	4490	4392	98	97.82%
阴沟肠杆菌	<i>Enterobacter cloacae</i> ASM298219v1	5023	4820	203	95.96%
化脓链球菌	<i>Streptococcus pyogenes</i> M1 GAS	1801	1639	162	91.00%

除嗜肺军团菌外，其他菌的滤过率仅达到了 90%，感染率最高的肺炎克雷伯氏菌更是达到了 98.87%，特征基因数目也较少，只有 61 个，达到了预期目标，对今后开展扩增实验有很大帮助。

3.2.2 普适性筛选结果

将 A 菌株在前两轮中获得的特征基因与这一菌种的所有菌株进行 BLAST 局部对比，剔除没有匹配序列的基因，即只在 A 菌株中存在而其他同种株不存在的基因，从而使特征基因具有最基本的普适性。此轮筛选还剔除了序列长度小于 400bp 的基因，有利于引物设计。筛选结果如表 7 所示。

表 7. 普适性筛选结果

菌种	代表性菌株	基因数量	剔除基因数	特征基因数	滤过率	第二轮特征基因数	较上一轮减少
肺炎克雷伯菌	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> HS11286	5404	5374	30	99.44%	61	31
鲍曼不动杆菌	<i>Acinetobacter baumannii</i> ASM74664v1	4043	4038	5	99.88%	253	248
大肠杆菌	<i>Escherichia coli</i> str. K-12 substr. MG1655	4566	4551	15	99.67%	197	182
肺炎链球菌	<i>Streptococcus pneumoniae</i> R6	1967	1950	17	99.14%	57	40
金黄色葡萄球菌	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> NCTC 8325	2872	2842	30	98.96%	43	13
嗜肺军团菌	<i>Legionella pneumophila</i> subsp. <i>pneumophila</i> str. Philadelphia 1	3119	3020	99	96.83%	531	432
流感嗜血杆菌	<i>Haemophilus influenzae</i> Rd KW20	1765	1737	28	98.41%	101	73
铜绿假单胞菌	<i>Pseudomonas aeruginosa</i> PAO1	5697	5691	6	99.89%	9	3
卡他莫拉菌	<i>Moraxella catarrhalis</i> BBH18	1738	1707	31	98.22%	109	78
嗜麦芽窄食单胞菌	<i>Stenotrophomonas maltophilia</i> K279a	4490	4460	30	99.33%	98	68
阴沟肠杆菌	<i>Enterobacter cloacae</i> ASM298219v1	5023	4921	102	97.97%	203	101
化脓链球菌	<i>Streptococcus pyogenes</i> M1 GAS	1801	1703	98	94.56%	162	64

此轮结果的特征基因除个例外较上一轮有约 40%减少，以肺炎克雷伯氏菌为例，经过第三轮筛选，特征基因数目减少了一半以上，只剩下 30 个。这 30 个特征基因具有很好的普适性，广泛存在于肺炎克雷伯氏菌的种系里。经过这一轮筛选得到的特征基因，可以有效提高病原菌检测的准确性，降低误判漏判的概率。

3.2.3 最优筛选结果

我们将第三轮获得的基因与细菌全基因组进行 BLAST 局部比对，通过分析比对结果，挑选了 3-5 个特异性最强、普适性最好的特征基因。表 8 是最终筛选结果。

表 8. 最优筛选结果

细菌名称	特征基因 locus tag	protein
铜绿假单胞菌	PA1369	hypothetical protein
	PA1371	hypothetical protein
	PA1935	hypothetical protein
嗜流感杆菌	HI0011	DNA polymerase III, psi subunit (holD)

	HI0436	competence protein D (comD)
	HI0940	predicted coding region HI0940
卡他莫拉菌	MCR_RS01105	hypothetical protein
	MCR_RS06295	lysozyme inhibitor
	MCR_RS06425	DUF2059 domain-containing protein
军团菌	CC46_RS0100955	HlyD family secretion protein
	CC46_RS0100965	FUSC family protein
	CC46_RS0106785	cyclase family protein
	CC46_RS0107645	NAD(P)-binding protein
金黄色葡萄球菌	V676_02278	phospholipase/carboxylesterase
	V676_02514	iron compound ABC transporter, iron compound-binding protein SirA
	V676_02537	hypothetical protein
化脓链球菌	SPy_0334	DUF177 domain-containing protein
	SPy_0407	hypothetical protein
	SPy_0428	ADP-ribosyl transferase
肺炎链球菌	SAMEA3431348_00219	glycosyltransferase involved exopolysaccharide (EPS) synthesis
	SAMEA3431348_00224	N-acetylglucosaminyl phosphatidylinositol deacetylase superfamily
	SAMEA3431348_02235	transposase, IS1380-SpnI related, truncation
肺炎克雷伯菌	KPHS_15450	putative arginine/ornithine antiporter
	KPHS_19280	putative DsbA oxidoreductase
	KPHS_31320	hypothetical protein
大肠杆菌	b2119	putative AAA(+) MoxR family ATPase YehL
	b2956	DUF1202 domain-containing protein YggM
	b3139	galactosamine-specific PTS enzyme IIC component
鲍曼不动杆菌	IX87_01165	methyltransferase type 11
	IX87_01545	hypothetical protein
	IX87_01565	hypothetical protein
阴沟肠杆菌	ECL_00924	LuxR family transcriptional regulator
	ECL_01364	putative acetyltransferase
	ECL_01719	hypothetical protein
	ECL_04449	hypothetical protein
	ECL_04191	hypothetical protein
嗜麦芽窄食单胞菌	SMLT_RS22540	RHS repeat-associated core domain-containing protein
	SMLT_RS22410	hypothetical protein
	SMLT_RS06670	ribonuclease inhibitor
	SMLT_RS09555	hypothetical protein
	SMLT_RS23230	hypothetical protein

我们的筛选标准是特异性由结果中杂菌数量最少决定，普适性由结果中同种菌株数量最多 (organisms) 决定。以肺炎克雷伯氏菌为例，我们从 30 个候选基因中挑选了 hit 数最多的三个基因，如表 9 所示，这三个基因的得分(score)都较高，而且杂菌数量较少，符合我们的筛选预期。

表 9. 肺炎克雷伯氏菌最优筛选记录

基因 locus tag	Organism	Number of Hits	Score
KPHS_19280	<i>Klebsiella pneumoniae</i>	87	1164
	<i>Klebsiella pneumoniae subsp. pneumoniae</i>	13	1164
KPHS_15450	<i>Klebsiella pneumoniae</i>	87	791

	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i>	13	791
	<i>Klebsiella pneumoniae</i>	83	1142
KPHS_31320	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i>	12	1142
	<i>Klebsiella pneumoniae</i> JM45	1	1142
	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> HS11286	1	1142

3.2.4 RAA 引物与探针的设计和筛选

我们根据得到的肺炎克雷伯氏菌特征序列设计了多个 RAA 引物，并进行了扩增效率测试，筛选出了扩增能力较好的引物和探针。引物的序列如表 10 所示，引物扩增实验和荧光探针筛选的结果如图 5 所示。由图 5a 可以看出，KPC2-5F-4R 和 KPC2-4F-5R 组合较 KPC2-8F-7R 有更好的准确性，而 KPC2-4F-5R 较 KPC2-5F-4R 纯净度更高。由图 5b 可以看出，探针 *exo-1* 较 *exo-2* 有更高的荧光强度，而 4F-5R+*exo-1* 较 5F-4R+*exo-2* 荧光强度更高。所以引物探针组合 4F-5R+*exo-1* 有很好的扩增效率，可以用于病原菌的基因检测。

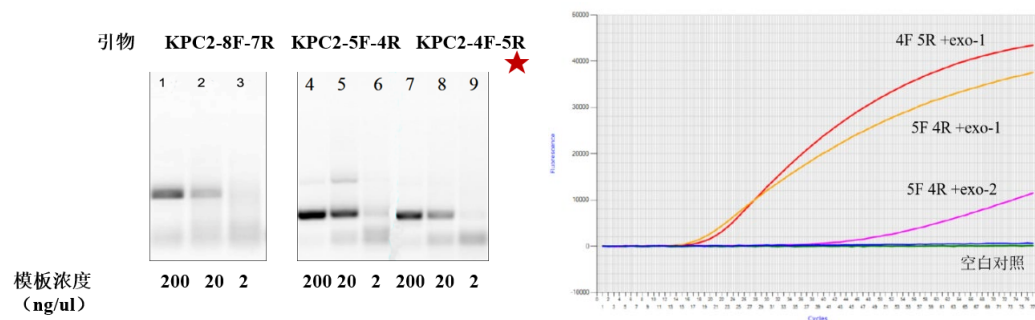


图 5 引物与荧光探针筛选结果

表 10 RAA 引物和探针序列

Primer/Probe	Sequence
KPC2-4-F	ACGCGATGGATACCGGCTCAGGCGCAACTG
KPC2-5-R	ACCACGGAACCAGCGCATTTTGGCGTAACG
KPC2-7-R	TAGCCGCAGCGGCGATGACGGCCTCGTGTGC
KPC2-8-F	ACTGACACTGGGCTCTGCACTGGCTGCGC
KPC2- <i>exo-1</i>	AGCTCATTCAAGGGCTTTCTTGCTGCCGCTGTGCTGGCTCGCAGCCAGC
KPC2- <i>exo-2</i>	AGTGTATGGCACGGCAAATGACTATGCCGTCGTCTGGCCCACTGGGGCG

3.3 特征引物筛选

我们根据设计的特征引物筛选算法（见图 2），并根据每一种菌的特点设定了不同筛选性能，筛选出了 12 种病原菌的候选引物，如表 11 所示。接着我们使用 BLAST 方法对得到的引物进行了分析，最终得到了肺炎克雷伯氏菌、大肠杆菌和铜绿假单胞菌的最优特异性引物对，如表 12 所示。

表 11 候选引物参数及结果统计

菌种	候选引物性能	候选引物数量
<i>Acinetobacter baumannii</i> 鲍曼不动杆菌	单序列普适度 > 60%，双引物对普适度 > 60%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	18 组
<i>Enterobacter cloacae</i> 阴沟肠杆菌	单序列普适度 > 60%，双引物对普适度 > 60%，全匹配特异度 = 100%，模糊匹配差异量 > 5bp	4 组

<i>Escherichia coli</i> 大肠杆菌	单序列普适度 > 60%，双引物对普适度 > 50%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	10 组
<i>Haemophilus influenzae</i> 流感嗜血杆菌	单序列普适度 > 60%，双引物对普适度 > 55%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	64 组
<i>Klebsiella pneumoniae</i> 肺炎克雷伯氏菌	单序列普适度 > 60%，双引物对普适度 > 60%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	1010 组
<i>Legionella pneumophila</i> 嗜肺军团菌	单序列普适度 > 60%，双引物对普适度 > 90%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	8 组
<i>Moraxella catarrhalis</i> 卡他莫拉菌	单序列普适度 > 90%，双引物对普适度 > 80%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	3245 组
<i>Pseudomonas aeruginosa</i> 铜绿假单胞菌	单序列普适度 > 60%，双引物对普适度 > 70%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	438 组
<i>Staphylococcus aureus</i> 金黄色葡萄球菌	单序列普适度 > 60%，双引物对普适度 > 70%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	290 组
<i>Stenotrophomonas maltophilia</i> 嗜麦芽窄食单胞菌	单序列普适度 > 60%，双引物对普适度 > 80%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	5 组
<i>Streptococcus pneumoniae</i> 肺炎链球菌	单序列普适度 > 60%，双引物对普适度 > 90%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	485 组
<i>Streptococcus pyogenes</i> 化脓链球菌	单序列普适度 > 60%，双引物对普适度 > 80%，全匹配特异度 = 100%，模糊匹配差异量 > 6bp	500 组

表 12 最优特征引物信息

菌种	引物名称	引物序列
<i>Klebsiella pneumoniae</i>	KP-hx02-F	GTATCGGTATTCTGGCTTAGCACAAAACC
	KP-hx02-R	GCGGATGAAGTAGGTCAGCGGTAAACAGGA
<i>Escherichia coli</i>	Ecoli-hx01-F	GGTAAAAGGAAACCAGGGCGGCTAAATAA
	Ecoli-hx01-R	TAAGTTCATCAGGGACATCGCAAACAATAT
<i>pseudomonas aeruginosa</i>	PA-hx01-F	AAGACATCCTGACACAAATCCGTGCGGCGG
	PA-hx01-R	GTTCTGCGTCGTGCTATTGACTTCATCGTC

3.4 特征引物实验验证

采用核酸扩增的方法进行菌种鉴定的关键在于进行核酸扩增时所用引物的特异性与普适性，其中最为重要的是引物的特异性。只有具有较高的种间特异性的引物才能确保在扩增时只对引物所对应的菌种 DNA 进行扩增而对其他菌种的 DNA 不扩增。为验证我们设计的引物具有良好的特异性，我们提取了肺炎克雷伯菌、大肠杆菌、铜绿假单胞菌、沙门氏菌和鲍曼不动杆菌等常见临床感染菌种的代表菌株的基因组以及经过事先验证的临床样本感染菌株的基因组，将我们设计的引物对这些基因组进行 PCR 扩增并对 PCR 扩增产物进行琼脂糖凝胶电泳，根据扩增以及电泳的结果对引物特异性进行验证，以下是验证结果。

针对我们所设计的肺炎克雷伯菌特异性引物，我们将其对包括两株肺炎克雷伯菌近缘菌株在内的 11 株代表菌株的基因组和 10 株已确认为肺炎克雷伯菌的临床样本的基因组进行 PCR 扩增，结果如图 6 所示。可见 11 株代表菌株基因组中只有肺炎克雷伯氏菌代表菌株基因组 PCR 扩增产物电泳后有条带，10 株已确认为肺炎克雷伯菌的临床样本基因组 PCR 扩增产物电泳后均有条带，且所有条带在凝胶上迁移距离大致相同，表明 PCR 扩增产物为长度相同的 DNA 片段。扩增及电泳结果表明我们所设计的肺炎克雷伯菌特异性引物具有很好的种内普适性与种间特异性，能够实现肺炎克雷伯菌基因组片段的特异性扩增。

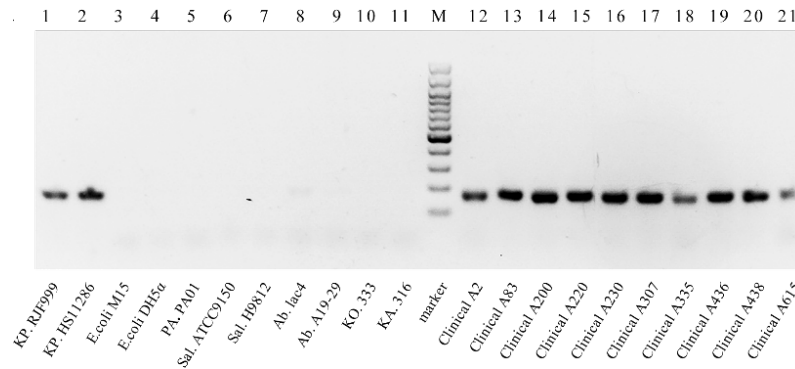


图 6 肺炎克雷伯菌特征引物对 11 株代表菌株和已确认为肺炎克雷伯菌的 10 株临床样本基因组 PCR 扩增结果 M,100bp marker; 1,*Klebsiella pneumoniae* RJF999; 2, *Klebsiella pneumoniae* HS11286; 3,*Escherichia coli* M15; 4, *Escherichia coli* DH5a; 5, *Pseudomonas aeruginosa* PAO1; 6, *Salmonella* ATCC 9150; 7, *Salmonella* H9812; 8, *Acinetobacter baumannii* lac4; 9, *Acinetobacter baumannii* A19-29; 10, *Klebsiella oxytoca* 333; 11, *Klebsiella aerogenes* 316. 12-21,不同的肺炎克雷伯氏菌临床菌株。12, *Klebsiella pneumoniae* A2; 13, *Klebsiella pneumoniae* A83; 14, *Klebsiella pneumoniae* A200; 15, *Klebsiella pneumoniae* A220; 16, *Klebsiella pneumoniae* A230; 17, *Klebsiella pneumoniae* A307; 18, *Klebsiella pneumoniae* A335; 19, *Klebsiella pneumoniae* A436; 20, *Klebsiella pneumoniae* A438; 21, *Klebsiella pneumoniae* A615.

针对我们设计的大肠杆菌特异性引物,我们将其对 9 株代表菌株的基因组和包含大肠杆菌在内的 24 株临床样本基因组进行 PCR 扩增,结果如图 7 及图 8 所示。可见 9 株代表菌株基因组中只有大肠杆菌对应基因组 PCR 扩增产物电泳后有条带,24 株临床样本基因组中 5 株已确认为大肠杆菌样本基因组扩增产物电泳后均有条带,其余样本基因组扩增产物电泳后则无条带。扩增结果表明我们所设计的大肠杆菌特异性引物具有良好的种间特异性与种内普适性,能够实现对大肠杆菌基因组片段的特异性扩增。

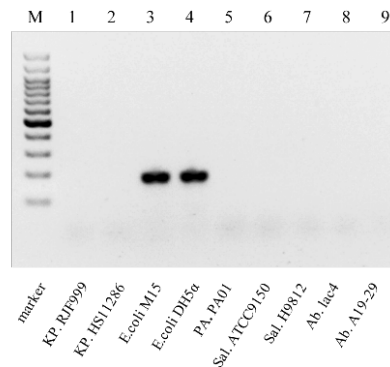


图 7 大肠杆菌特征引物对 9 株代表菌株基因组 PCR 扩增结果 M,100bp marker; 1,*Klebsiella pneumoniae* RJF999; 2, *Klebsiella pneumoniae* HS11286; 3,*Escherichia coli* M15; 4, *Escherichia coli* DH5a; 5, *Pseudomonas aeruginosa* PAO1; 6, *Salmonella* ATCC 9150; 7, *Salmonella* H9812; 8, *Acinetobacter baumannii* lac4; 9, *Acinetobacter baumannii* A19-29.

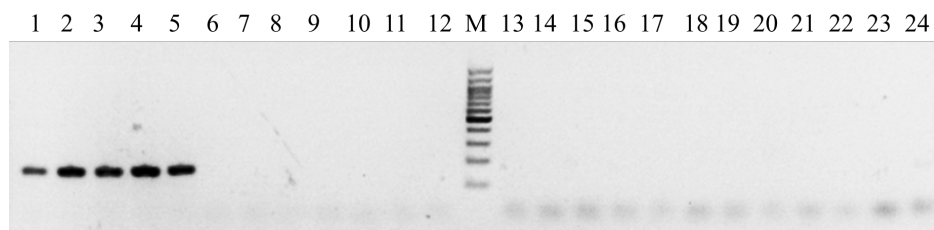


图 8 大肠杆菌特征引物对 24 株临床样本基因组 PCR 扩增结果 M, 100bp Marker; 1, *Escherichia coli* 158; 2, *Escherichia coli* 159; 3, *Escherichia coli* 160; 4, *Escherichia coli* 161; 5, *Escherichia coli* 162; 6, *Enterobacter cloacae* 37; 7, *Acinetobacter baumannii* 458; 8, *Acinetobacter baumannii* 460; 9, *Acinetobacter baumannii* 466; 10, *Acinetobacter baumannii* 476; 11, *Acinetobacter baumannii* 477; 12, *Salmonella enterica* 113; 13, *Salmonella enterica* 124; 14, *Salmonella enterica* 173; 15, *Salmonella enterica* 233; 16, *Salmonella enterica* 252; 17, *Pseudomonas aeruginosa* 124; 18, *Pseudomonas aeruginosa* 131; 19, *Pseudomonas aeruginosa* 153; 20, *Staphylococcus aureus* 224; 21, *Staphylococcus aureus* 244; 22, *Staphylococcus aureus* 252; 23, *Staphylococcus aureus* 600; 24, *Staphylococcus aureus* 602.

针对我们设计的铜绿假单胞菌特异性引物，我们将其对 9 株代表菌株的基因组和 24 株包括铜绿假单胞菌在内的已经过验证的临床样本基因组进行 PCR 扩增，结果如图 9 与图 10 所示。可见 9 株代表菌株基因组中只有铜绿假单胞菌代表菌株基因组 PCR 扩增产物电泳后有条带，24 株已经过验证的临床样本基因组中 3 株确认为铜绿假单胞菌的临床样本基因组 PCR 扩增产物电泳后均有条带，其余基因组扩增产物电泳则无条带。扩增结果表明我们所设计的铜绿假单胞菌特异性引物具有良好的种间特异性与种内普适性。

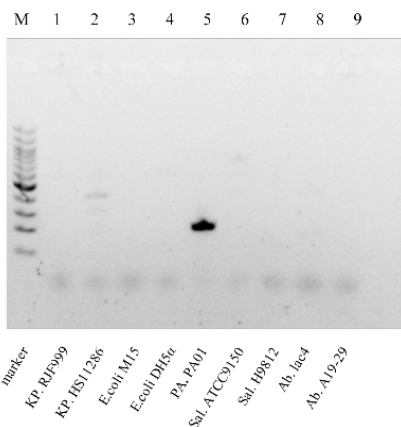


图 9 铜绿假单胞菌特征引物对 9 株代表菌株基因组 PCR 扩增结果 M,100bp marker; 1,*Klebsiella pneumoniae* RJF999; 2, *Klebsiella pneumoniae* HS11286; 3,*Escherichia coli* M15; 4, *Escherichia coli* DH5 α ; 5, *Pseudomonas aeruginosa* PAO1; 6, *Salmonella* ATCC 9150; 7, *Salmonella* H9812; 8, *Acinetobacter baumannii* lac4; 9, *Acinetobacter baumannii* A19-29.

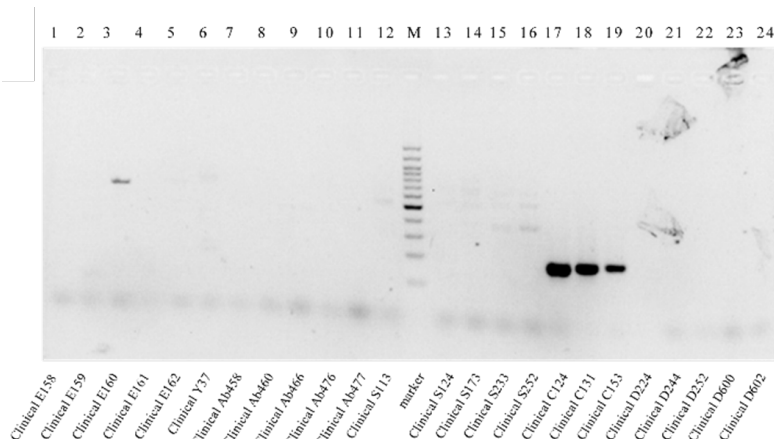


图 10 铜绿假单胞菌特征引物对 24 株临床样本基因组 PCR 扩增结果 M, 100bp Marker; 1, *Escherichia coli* 158; 2, *Escherichia coli* 159; 3, *Escherichia coli* 160; 4, *Escherichia coli* 161; 5, *Escherichia coli* 162; 6, *Enterobacter cloacae* 37; 7, *Acinetobacter baumannii* 458; 8, *Acinetobacter baumannii* 460; 9, *Acinetobacter baumannii* 466; 10, *Acinetobacter baumannii* 476; 11, *Acinetobacter baumannii* 477; 12, *Salmonella enterica* 113; 13, *Salmonella enterica* 124; 14, *Salmonella enterica* 173; 15, *Salmonella enterica* 233; 16, *Salmonella enterica* 252; 17, *Pseudomonas aeruginosa* 124; 18, *Pseudomonas aeruginosa* 131; 19, *Pseudomonas aeruginosa* 153; 20, *Staphylococcus aureus* 224; 21, *Staphylococcus aureus* 244; 22, *Staphylococcus aureus* 252; 23, *Staphylococcus aureus* 600; 24, *Staphylococcus aureus* 602.

针对我们所设计的鲍曼不动杆菌特征引物，我们将其对 9 株不同菌株的鲍曼不动杆菌基因组进行 PCR 扩增，结果如图 11 所示。可见在 9 株鲍曼不动杆菌中除 lac4 菌株外，其他 8 株菌株基因组 PCR 扩增产物电泳后均有条带。扩增结果表明我们所设计的鲍曼不动杆菌特征引物并不具备对鲍曼不动杆菌各菌株普适性扩增的能力，因此不适宜用于鲍曼不动杆菌的临床检测中。

M 1 2 3 4 5 6 7 8 9

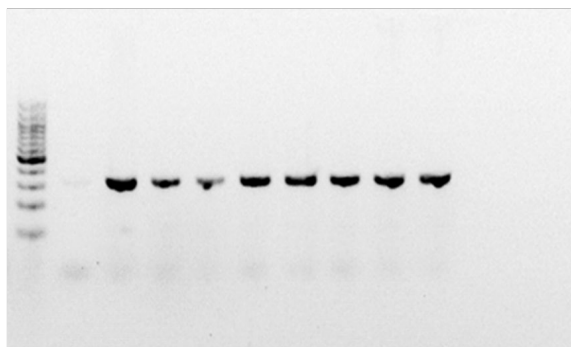


图 11 鲍曼不动杆菌特征引物对 9 株不同鲍曼不动杆菌菌株基因组扩增结果 M, 100bp Marker; 1, *Acinetobacter baumannii* lac4; 2, *Acinetobacter baumannii* A19-29; 3, *Acinetobacter baumannii* 458; 4, *Acinetobacter baumannii* 460; 5, *Acinetobacter baumannii* 466; 6, *Acinetobacter baumannii* 476; 7, *Acinetobacter baumannii* 477; 8, *Acinetobacter baumannii* ATCC19606; 9, *Acinetobacter baumannii* BNCC194496.

以上的实验结果均表明采用我们的方法设计出的特异性引物除鲍曼不动杆菌外均具备良好的种间特异性与种内普适性，能够实现对目标菌种基因组片段的特异性扩增，能够通过核酸体外扩增技术实现对目标菌种的特异性检测。同时，采用我们筛选出的特异性引物通过核酸体外扩增方法检测临床感染样本经平板稀释涂布法分离纯化所得单菌落的基因组的检测结果与医院事先对这些样本的检测结果相同，能够检测出目标菌种，表明我们所构建的特异性引物挖掘机制结合核酸体外扩增技术能够用于临床常见病原菌的检测。

3.2 采用 RAA 技术进行扩增检测的灵敏度检验

除利用 PCR 扩增和琼脂糖凝胶电泳验证我们得到的特异性引物是否具备良好的种间特异性外，本研究还采用 RAA 技术成功就肺炎克雷伯氏菌的常见耐药基因 KPC-2 完成了检测，检测结果十分成功，证明了利用 RAA 技术进行核酸体外扩增用于临床常见病原菌种类及耐药性快速检测的可行性。此外，本研究还对采用 RAA 技术进行扩增检测的灵敏度进行了检验。

在该实验中我们采用前期已证明扩增检测效果较好的 RAA 引物 KPC2-4F-5R 与荧光探针 KPC2-exo-1 组合，对不同浓度的模板进行扩增，检测扩增过程中的荧光强度，结果如图 12 与图 14 所示。可见采用 RAA 技术为基础的结合荧光探针的目标基因检测技术能够检测到浓度最低为 200fg/μl 的目标基因，灵敏度很高，具有很大的应用前景。

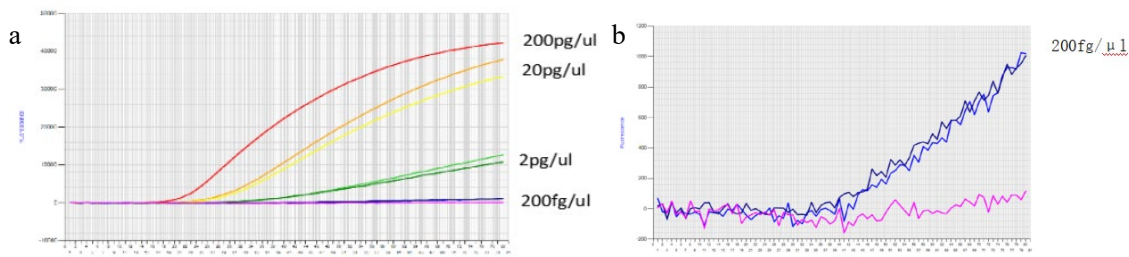


图 12 RAA 引物 KPC2-4F-5R 与荧光探针 KPC2-exo-1 组合检测不同浓度的 KPC-2 基因 (a 常规型, b 增强型)

4. 讨论

4.1 针对特征基因与特征引物筛选方法进行讨论:

在本课题项目的研究过程中 BLAST 的使用是极为频繁的。本研究利用 BLAST 工具实现对特征基因的挖掘和对我们初步挖掘出的多条特征引物进行进一步筛选。而在使用过程中根据实际的数对比情况合理地调整相应参数则是在使用数据分析软件过程的重中之重。

在对特征基因进行第一轮和第二轮的挖掘过程中我们需要将期望阈值 (Expect threshold, 即

BLAST 工具中的 e value) 调小, 一般是 10^{-5} 。由于该对比分析软件采用的是随机匹配出现序列的比对得分期望数, 其数值越低, 匹配到相同基因的概率就越低, 最终得到的序列就越少, 其特异性也越高。基于前两次的筛选过程我们需要特异性高的基因序列, 因此在这一部分所采用的期望阈值需要调小。而经过了两次筛选之后所得的基因数目为原基因数目的 1% 左右, 在接下来的筛选过程中需要得到普适性良好的基因, 因此需要调整期望阈值到 1000 左右, 以尽可能得到相似的匹配序列, 即代表了该特征基因的普适性较高。此外, 在第四轮对细菌的全基因组序列进行比对的过程中, 可以允许比对结果的 **organism** 中有少量杂菌存在, 因为此时经过三轮筛选能匹配上的 **organism** 数目已经较少, 只要这些匹配上的杂菌基本不会出现在人的呼吸道感染中即可。

在对针对目标菌种所设计的特征引物进行筛选时将待筛选引物对细菌全基因组进行 BLAST 比对时, 通过比较比对结果 **organism** 中目标菌种占有所有比对上的菌种数的比例确定最佳的特征引物。此时应注意比对结果中除了目标菌种外还有许多杂菌, 只要这些匹配上的杂菌基本不会出现在人的呼吸道感染中即可, 这样做是因为在目标菌种中普适性较好的特征引物往往会在 BLAST 比对中与目标菌种的近缘菌种发生匹配, 只要匹配上数量较多的几种近缘菌不出现在人呼吸道感染中即不会影响最终的检测结果, 在确保引物特异性的同时最大程度地提升引物的种内普适性, 避免假阴性的发生。

4.2 针对特征基因与特征引物分析结果的讨论

本研究利用 PCR 扩增和琼脂糖凝胶电泳验证我们得到的特异性引物的实验结果表明, 我们所设计的特异性引物除鲍曼不动杆菌外均具备良好的种内普适性与种间特异性, 表明我们所构建的特征引物挖掘方法是有效的。同时本研究实现了对临床样本的准确检测, 证明本研究所构建的特异性引物挖掘机制结合核酸体外扩增技术能够实现对临床常见病原菌的检测。^[9]

除此之外, 本研究略有不足的是在验证特征引物普适性与特异性的实验中所用临床样本基因组均来自于临床样本采用平板稀释涂布法分离纯化后所得单菌落提取的基因组, 未能实现对临床样本的直接检测, 还需在基因组提取的效率方面进行提高以应用于实际情况下的临床快速检测。^[10]

5. 结论

本课题首先通过对呼吸道常见耐药病原菌耐药基因的多次有效快速的筛查和分析得到了足够精准有效的特征基因。即, 对十二种呼吸道标本分离菌的主要菌株 (临床感染率达 90% 以上), 共计 42485 个基因组序列, 筛选的得到了 43 个特征基因。同时, 在筛查基因的过程中建立起来的病原菌特征基因快速准确的筛查体系, 能够快速实现菌株特征基因序列的筛查同时确保筛选基因的特异性和普适性。

其次, 利用大数据组学分析, 开发得到了相应的数据算法, 采用基因序列格式化和 N 元切分, 在多轮单双序列全匹配普适度和特异性过滤的作用下有效降低了目标量级。在添加了布隆过滤器和模糊算法等功能的作用下, 通过优化筛选路径, 有效提高了运行速度并得到了肺炎克雷伯氏菌、大肠埃希菌和铜绿假单胞菌的兼具特异性和普适度的特征引物对共 3 对 6 条。

通过单菌据、混合菌株和临床试验化的验证, PCR 的扩增试验数据均显示出该方法得到的特征性引物具有极佳的特异性, 能够快速实现对目标菌株的特异性扩增和精准鉴定。

6. 展望

相对来说, 本课题最为核心和关键的是建立了一种 BLAST、数据分析与大数据筛选相结合的研究方法, 能够针对临床常见病原菌进行快速的检查和分析, 提供最佳的基因靶向的位点和序列以有助于此课题后续的发展。当然, 基因序列的筛选方式和方法也有待进一步的优化和解决, 目前该课题主要针对的是已广泛接受和认知的病原菌, 且运用和考虑的问题关键点尚不包括基因突变和代偿性功能转移等一系列实际可能存在的关键性问题。在动态变化的细菌感染过程中, 如何建立一个动态, 更加符合实际需求和生产生活需要的筛选方法和路径有待后续的分析和解决。

在后续的课题发展中, 需要实现该功能靶向序列的医学转化, 即利用该方式方法来完成从基因序列到临床医疗表征与识别的转化。在耐药革兰氏阳性菌感染的快速诊断和智能化治疗方案的

这一科技部政府间重大专项成果转化研究成果中,本课题的研究进展能够为后续临床常见耐药病原菌快速检测试剂盒的研发提供最为基本的序列支撑,但在临床转化前,依旧有着很多亟待解决的科学问题,而这些科学问题的解决也势必会对世界耐药菌的检测和治疗格局带来新的突破和发展!

7. 参考文献

- [1] Martens E, Demain A L. The antibiotic resistance crisis, with a focus on the United States[J]. *Journal of Antibiotics*, 2017, 70(5): 520-526.
- [2] Rice L B. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESCAPE[J]. *Journal of Infectious Diseases*, 2008, 197(8): 1079-1081.
- [3] Dodds D R. Antibiotic resistance: A current epilogue[J]. *Biochemical Pharmacology*, 2017, 134: 139-146.
- [4] Carlet J, Collignon P, Goldmann D, et al. Society's failure to protect a precious resource: antibiotics[J]. *Lancet*, 2011, 378(9788): 369-371.
- [5] Varadi L, Luo J L, Hibbs D E, et al. Methods for the detection and identification of pathogenic bacteria: past, present, and future[J]. *Chemical Society Reviews*, 2017, 46(16): 4818-4832.
- [6] Altun O, Almuhayawi M, Ullberg M, et al. Clinical Evaluation of the FilmArray Blood Culture Identification Panel in Identification of Bacteria and Yeasts from Positive Blood Culture Bottles[J]. *Journal of Clinical Microbiology*, 2013, 51(12): 4130-4136.
- [7] 胡付品, 郭燕, 朱德妹, et al. 2017年CHINET中国细菌耐药性监测[J]. *中国感染与化疗杂志*, 2018, 18(03): 241-251.
- [8] 杨霞, 陈陆, 王川庆. 16S rRNA基因序列分析技术在细菌分类中应用的研究进展[J]. *西北农林科技大学学报(自然科学版)*, 2008(02): 55-60.
- [9] 吕蓓, 程海荣, 严庆丰, et al. 体外核酸快速扩增技术的发展和不断创新[J]. *中国生物工程杂志*, 2011, 31(03): 91-96.
- [10] Thoendel M, Jeraldo P R, Greenwood-Quaintance K E, et al. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing[J]. *Journal of Microbiological Methods*, 2016, 127: 141-145.

8. 致谢

本项目是在生命科学技术学院马伟老师的悉心指导下完成的。马伟老师渊博的专业知识,严谨的治学态度,精益求精的工作作风让我们受益匪浅。在此,谨向马伟老师表示崇高的敬意和衷心的感谢!

本项目的顺利完成离不开实验室师兄师姐的帮助。马伟老师课题组的薛松师兄热情指导我们生信分析、实验部分的有关操作。其力格尔师姐、岳烨师姐等也在实验操作上热心帮助我们。在此向各位师兄师姐表示感谢!

本项目的相关程序设计与上海画心智能科技有限公司共同完成,在此感谢他们的付出。

最后感谢上海交通大学大学生创新实践计划(IPP20108)对本项目的经费和平台支持,使得我们能够在大学二年级就能够深入国家重点实验室进行科研学习。