

計畫編號：MOHW109-CDC-C-315-144416

衛生福利部疾病管制署 109 年署內科技研究計畫

計畫名稱：高通量定序(NGS)平台開發與應用

## 109 年 度 研 究 報 告

執行單位：檢驗及疫苗研製中心

計畫主持人：慕蓉蓉

協同主持人：

研究人員：林鈺棋

研究人員：陳筱蓉

執行期間：109 年 1 月 1 日至 109 年 12 月 31 日

## 目錄

	頁	碼
目錄		1
計畫中文摘要		2
計畫英文摘要		3
計畫內容		
一、前言		4
二、材料與方法		6
三、結果		8
四、討論		11
五、結論與建議		13
六、參考資料		14
六、圖、表		16

共 (21) 頁

## 計畫中文摘要：

高通量定序（high-throughput sequencing），又稱為下世代定序（Next Generation Sequencing, NGS）具有龐大數據之產出以及 de novo 定序分析的強大能力，有別於以往僅能對已知病原體基因序列進行偵測，近年來廣泛運用在傳染病原的偵測與追蹤，將分子診斷提升至全基因領域。目前 NGS 可應用於不明原因傳染病原偵測、新興微生物（病原）之鑑定分析、微生物全基因體定序、微生物分型、傳染病群聚調查、抗藥性基因分析比對以及微生物菌叢（microbiome）的變化追蹤等。

本計畫完成建置以第二代搭配第三代 NGS 全基因定序技術進行細菌全基因體組裝、基因分析以及比對，期望運用高通量定序偵測平台建立病原體基因序列組裝分析的流程，適當串連各種分子檢驗技術之優點，以期能於第一時間將罕見或變異之病原體全基因解密，強化防疫時效與避免不明傳染病對社會造成衝擊。加強病原體基因序列資料之彙整分析，瞭解其可能的感染源、疾病的流行趨勢及協助疫情調查外，還可獲取更多生物資訊以增進對該病原的瞭解及認識，作為未來防疫政策擬訂及相關疾病研究的重要參考。

關鍵詞：高通量，次世代定序

計畫英文摘要：

keywords : high-throughput, next generation sequencing

High-throughput sequencing, also referred to as next generation sequencing has been proved itself to be robust and expanded to the whole genome level. The NGS has dramatically changed the molecular diagnostic arena in clinical microbiology and applied to detection of unknown disease associated pathogens, genotypic resistance testing, outbreak investigation and investigation of microbial population diversity.

In this study, miseq and PacBio NGS platforms have been applied in the genome assembly and comparison. We expect to establish pathogen genome analysis platform by adequately integrating many kinds of molecular diagnostic techniques with NGS system. Through the analysis of pathogen sequences and related epidemiological information, we can assist the clarification of infectious pathogens and infection routes, provide a reference for disease prevention policy assessment and formulating monitor the drug resistance of pathogens and even provide a foundation for future developing diagnostic techniques.

## 本文

### 一、前言：

高通量定序又被稱為 deep sequencing, 下世代定序 (next generation sequencing, NGS), 其成功超越傳統 sanger sequencing, 不但可以一次產出 Gb 至 Tb 大量之序列資料, 且無需事先設計引子或探針, 直接針對未知基因進行序列分析, 再加上 de novo 定序組裝的強大能力, 可針對新興未知病原基因體完全解密[1]。

近年來高通量定序廣泛應用在傳染病原的偵測與追蹤, 例如不明原因傳染病原偵測、新興微生物 (病原) 之鑑定分析、微生物全基因體定序、微生物分型、傳染病群聚調查、抗藥性基因分析比對以及微生物菌叢 (microbiome) 的變化追蹤等[2-9]。

目前市面上有數種高通量定序平台針對不同目的, 數據資料的產出量, 每筆序列的長度及組裝功能的差異, 可供選擇使用。高通量定序在診斷應用上可分成兩種策略: whole exome sequencing (WES) 以及 whole genome sequencing (WGS) [10]。前者利用已知基因的序列, 大量篩選可能之變異 (例如抗藥變異) [11], 或是追蹤菌叢的變化 (例如以 16S 基因分析細菌種類) [12]。後者則利用 de novo sequencing 的方式, 可以組裝全基因體序列 (2011 年德國新型出血性大腸桿菌 E. Coli O104), 或者未知病原的偵測以及

群聚事件的調查[4-6]。

高通量序列分析平台彌補傳統的病原體檢測技術僅能對已知病原體基因序列進行檢測，且單次的試驗裡只能檢測一個或數種已知的病原體的不足，且具有大規模 de novo 定序分析的強大能力，提高病原全基因解析的能力與速度，隨著基因定序方法的進步，能迅速累積病原體基因序列，得以全面性分析病原體之基因體結構。因此，本計畫將選取不同之高通量定序平台分別運用於未知病原之偵測、全基因體定序組裝以及群聚事件的調查等。未來面對未知的新興傳染病時，可以此系統進行即時偵測及基因解密；且針對發生群聚時，進行分析比對尋找可能知感染來源，以期能即時發現不明與罕見的病原體，快速找出感染源，強化防疫時效與避免不明傳染病對社會造成衝擊。

## 二、材料與方法

### 1、實驗菌株

標準菌株使用包含 20 個菌種之混樣核酸(ATCC MSA-1003)，測試樣品使用包含 8 種細菌及 2 種真菌之 ZymoBIOMICS 混樣核酸，測試臨床菌種則挑選包括 *Corynebacterium striatum*、*Acinetobacter baumannii*、*Enterobacter asburiae*、*Borrelia burgdorferi* 和 *Leptospira borgpetersenii* 等 5 種菌種之核酸。

### 2、16S rRNA 總體基因體學之基因庫建置及 NGS 定序

參照 illumina 16S rRNA 總體基因體學之標準流程，針對 16S rRNA V3-V4 高變異區設計引子，再利用 PCR 方式合成，引子序列如下：

(1)16S Amplicon PCR Forward Primer：5'-TCGTCGGCAGCGTCA  
GATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3'

(2)16S Amplicon PCR Reverse Primer：5'- GTCTCGTGGGCTCGGA  
GATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

長片段 16S rRNA 總體基因體學採用通用型引子 8F 及 1492R。

反應試劑使用 2x KAPA HiFi HotStart ReadyMix(Roche)，取 1 $\mu$ L 樣品核酸與 1X KAPA HiFi HotStart ReadyMix、200nM forward primer 及 200nM reverse primer 混合，混合物以 Biometra 系統

(Biometra, Analytik Jena)進行反應，反應條件為 95°C作用 3 分鐘，再進行 25 次循環反應 (95°C作用 30 秒鐘，55°C作用 30 秒鐘，72°C作用 30 秒鐘)，最後 72°C作用 5 分鐘後，維持在 4°C保存。

將上述 PCR 產物經 AMPure XP beads(Beckman Coulter)純化後，進行 Index PCR，反應試劑使用 2x KAPA HiFi HotStart ReadyMix(Roche)，取 5 $\mu$ LPCR 產物與 1X KAPA HiFi HotStart ReadyMix、5 $\mu$ L Nextera XT Index Primer 1 及 Primer 2 混合，混合物以 Biometra 系統 (Biometra, Analytik Jena)進行反應，反應條件為 95°C作用 3 分鐘，再進行 8 次循環反應 (95°C作用 30 秒鐘，55°C作用 30 秒鐘，72°C作用 30 秒鐘)，最後 72°C作用 5 分鐘後，維持在 4°C保存，上述 Index PCR 產物經 AMPure XP beads(Beckman Coulter)純化後，以 Illumina iSeq 定序。

### 3、Bioinformatics 分析

NGS 定序資料以 Local Run Manager(illumina)之 16S Metagenomics Analysis Module 進行分析，比對資料庫使用 Greengene database[13]。



### 三、 結果

#### 1、 建立 16S rRNA V3-V4 高變異區域總體基因體學分析技術

為運用高通量定序(NGS)於16S rRNA總體基因體學分析，遴選包含20個菌種之混樣核酸(ATCC MSA-1003)作為標準品，利用PCR合成16S rRNA V3-V4高變異區域片段後，經illumina iSeq進行NGS定序，再以illumina定序儀內建16S Metagenomics App分析(圖一)。結果顯示本計畫16S rRNA V3-V4區域總體基因體學分析結果，與illumina公司對於相同標準品ATCC MSA-1003之分析結果一致，皆可區分標準品中不同菌屬，顯示本計畫已完成建立16S rRNA V3-V4區域總體基因體學分析技術。

#### 2、 分析測試樣品及臨床菌株

進一步挑選包含 8 種細菌及 2 種真菌之 ZymoBIOMICS 混樣核酸作為測試樣品，其中包含 *Pseudomonas aeruginosa*、*Escherichia coli*、*Salmonella enterica*、*Lactobacillus fermentum*、*Enterococcus faecalis*、*Staphylococcus aureus*、*Listeria monocytogenes*、*Bacillus subtilis* 等 8 種細菌及 *Saccharomyces cerevisiae* 與 *Cryptococcus neoformans* 2 種真菌 (表一)。結果顯示，以 16S rRNA V3-V4 區域總體基因體學分析測試樣品，可偵測出包括 *Pseudomonas*、*Escherichia*、*Lactobacillus*、*Enterococcus*、*Staphylococcus*、*Listeria*、

*Bacillus* 等 7 種細菌，但僅能分析至屬名(genus)，無法分析至種名(species)。其中無法鑑定 *Salmonella* 菌屬，卻額外鑑定出測試樣品未包含的 *Enterobacter* 菌屬 (紅色標示)。2 種真菌則無法以 16S 鑑定。

經比對序列發現，*S. enterica* 與 *E. cloacae* 之 PCR 產物前端 150bp 與末端 150bp 序列皆一致，顯示兩者無法區分(圖二)，參照 16S Metagenomics 軟體之分析邏輯，因而將 *Salmonella* 菌屬鑑定為 *Enterobacter* 菌屬。

接續分別以 *Corynebacterium striatum*、*Acinetobacter baumannii*、*Enterobacter asburiae*、*Borrelia burgdorferi* 和 *Leptospira borgpetersenii* 菌種測試 16S rRNA V3-V4 區域之解析能力(表二)，結果顯示，16S rRNA V3-V4 區域總體基因體學技術可正確鑑定出這 5 種菌株之屬名 *Corynebacterium*、*Acinetobacter*、*Enterobacter*、*Borrelia* 和 *Leptospira* 等菌屬。

### 3、發展長片段 16S rRNA 之總體基因體學分析技術

由於 16S rRNA V3-V4 區域總體基因體學技術受限於片段長度及鑑別能力之限制，為更有效鑑定菌種，故進一步發展長片段 16S rRNA 總體基因體學分析技術。

分別以ATCC MSA-1003及ZymoBIOMICS混樣核酸作為測試樣品，使用16S rRNA通用型引子8F及1492R(I)，以PCR合成長片段16S rRNA後，利用NGS定序，再以16S Metagenomics App分析，結果如表三。結果顯示利用長片段16S rRNA總體基因體學分析，可將兩項測試樣品內的混樣核酸正確區分至種名，然而ATCC MSA-1003混樣核酸中的18% *E. coli* 菌種分別鑑定為13.2% *E. albertii*及1.5% *E. coli*；ZymoBIOMICS混樣核酸中的12% *E. coli* 菌種分別鑑定為7.2% *E. albertii*及0.7% *E. coli*，顯示兩項測試樣品內的*E. coli*核酸大部分鑑定為*E. albertii*。

#### 四、討論

本計畫完成建立 16S rRNA V3-V4 高變異區域總體基因體學技術，透過 ZymoBIOMICS 混樣核酸鑑定出該核酸未包含的 *Enterobacter* 菌屬之意外結果，進一步研究分析如下：

1、16S rRNA V3-V4 區域之 PCR 產物長度為 465bp，大於 iSeq 讀取長度 150 bp，故僅能讀取 PCR 產物前端 150bp 與末端 150bp 序列，無法讀取中段序列。

2、經比對序列發現，*S. enterica* 與 *E. cloacae* 之 PCR 產物前端 150bp 與末端 150bp 序列皆一致，顯示兩者無法區分(圖二)，參照 16S Metagenomics 軟體之分析邏輯，卻將 *Salmonella* 菌屬鑑定為 *Enterobacter* 菌屬。

綜上，利用 16S V3-V4 rRNA 總體基因體學分析，受限於 illumina iSeq 單一定序長度上限為 150bp，僅能鑑定至菌屬，且可能發生鑑定不正確之結果。因此，本計畫進一步發展長片段 16S rRNA 之總體基因體學技術，可鑑定細菌至種名(species)，然而依測試樣品之鑑定結果，大部分 *E. coli* 鑑定為 *E. albertii* 菌種，僅有少部分鑑定為 *E. coli*，此情況已有研究報導，使用 16S rRNA 全長序列作為菌種鑑定之標準，無法有效區分 *E. coli*、*E. fergusonii* 及 *E. albertii* 等三菌種，顯示 16S rRNA 序列不適用於鑑別關係非常相近之菌種[14, 15]，但對於大多數

菌種，16S rRNA 序列仍具有相當的鑑別力。

## 五、 結論與建議

1、 本計畫以建置 NGS 平台以應用於傳染病為主軸，歷年成果摘述如下：

(1) 第一(106)年完成 NGS 平台於不明原因感染原之檢驗應用，並制訂不明原因知感染源之檢驗流程。

(2) 第二(107)年運用兩種 NGS 平台（短片段搭配長片段組合校正），完成全基因組裝定序，並進行解密分析比對。

(3) 第三(108)年利用 NGS 產出大量序列，進行重要抗藥基因 NDM-5 群聚事件之追蹤分析。

2、 本計畫為第四(109)年計畫，透過建立 16S rRNA V3-V4 高變異區域及長片段 16S rRNA 之總體基因體學技術，可進行混樣菌屬甚至菌種之鑑定，隨著大數據科技的發展，未來 16S 總體基因體學分析應用於微生物菌叢(microbiota)之精準醫療，將扮演重要角色。此外，本計畫亦同時建立 Target NGS 之標準建庫方法，後續將運用 amplicon based target enrichment 兼具高靈敏度及多標的偵測之優勢，持續精進 NGS 檢測技術。

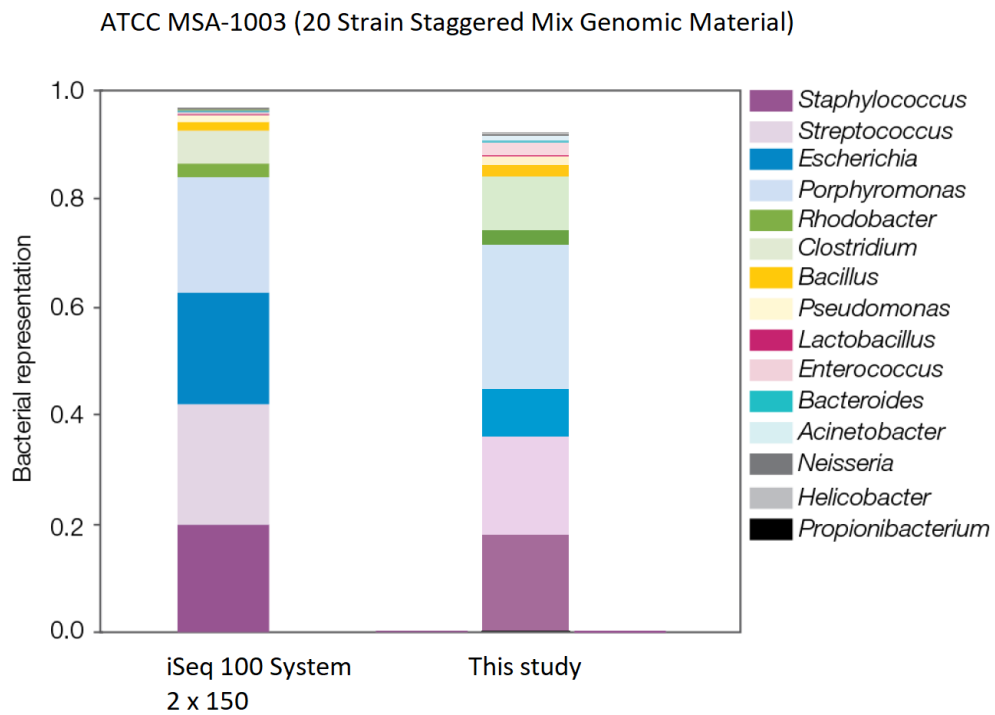
## 六、 參考資料

- [1] Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-Generation Sequencing for Infectious Disease Diagnosis and Management: A Report of the Association for Molecular Pathology. *The Journal of molecular diagnostics : JMD*. 2015;17:623-34.
- [2] Monira S, Nakamura S, Gotoh K, Izutsu K, Watanabe H, Alam NH, et al. Metagenomic profile of gut microbiota in children during cholera and recovery. *Gut pathogens*. 2013;5:1.
- [3] Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS pathogens*. 2012;8:e1002824.
- [4] Deng X, den Bakker HC, Hendriksen RS. Genomic Epidemiology: Whole-Genome-Sequencing-Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens. *Annual review of food science and technology*. 2016;7:353-74.
- [5] Brzuszkiewicz E, Thurmer A, Schuldes J, Leimbach A, Liesegang H, Meyer FD, et al. Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Archives of microbiology*. 2011;193:883-91.
- [6] Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PloS one*. 2009;4:e4219.
- [7] Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. *The New England journal of medicine*. 2008;358:991-8.
- [8] Clausen PT, Zankari E, Aarestrup FM, Lund O. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *The Journal of antimicrobial chemotherapy*. 2016.
- [9] Cannas A, Mazzarelli A, Di Caro A, Delogu G, Girardi E. Molecular Typing of *Mycobacterium Tuberculosis* Strains: A Fundamental Tool for Tuberculosis Control and Elimination. *Infectious disease reports*. 2016;8:6567.
- [10] Lapin V, Mighion LC, da Silva CP, Cuperus Y, Bean LJ, Hegde MR. Regulating whole exome sequencing as a diagnostic test. *Human genetics*. 2016;135:655-73.
- [11] Fisher RG, Smith DM, Murrell B, Slabbert R, Kirby BM, Edson C, et al. Next generation sequencing improves detection of drug resistance mutations in infants after PMTCT failure. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2015;62:48-53.
- [12] Allen HK, Bayles DO, Looft T, Trachsel J, Bass BE, Alt DP, et al. Pipeline for amplifying and analyzing amplicons of the V1-V3 region of the 16S rRNA gene. *BMC research notes*. 2016;9:380.

- [13] 16S Metagenomic Sequencing Library Preparation. Available at: [https://support.illumina.com/content/dam/illumina-support/documents/documentation/software\\_documentation/local-run-manager/local-run-manager-16S-metagenomics-workflow-guide-1000000057866-00.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/local-run-manager/local-run-manager-16S-metagenomics-workflow-guide-1000000057866-00.pdf).
- [14] Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced *Escherichia coli* genomes. *Microbial ecology*. 2010;60:708-20.
- [15] Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen KY. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infec*. 2008;14:908-34.



## 七、圖、表

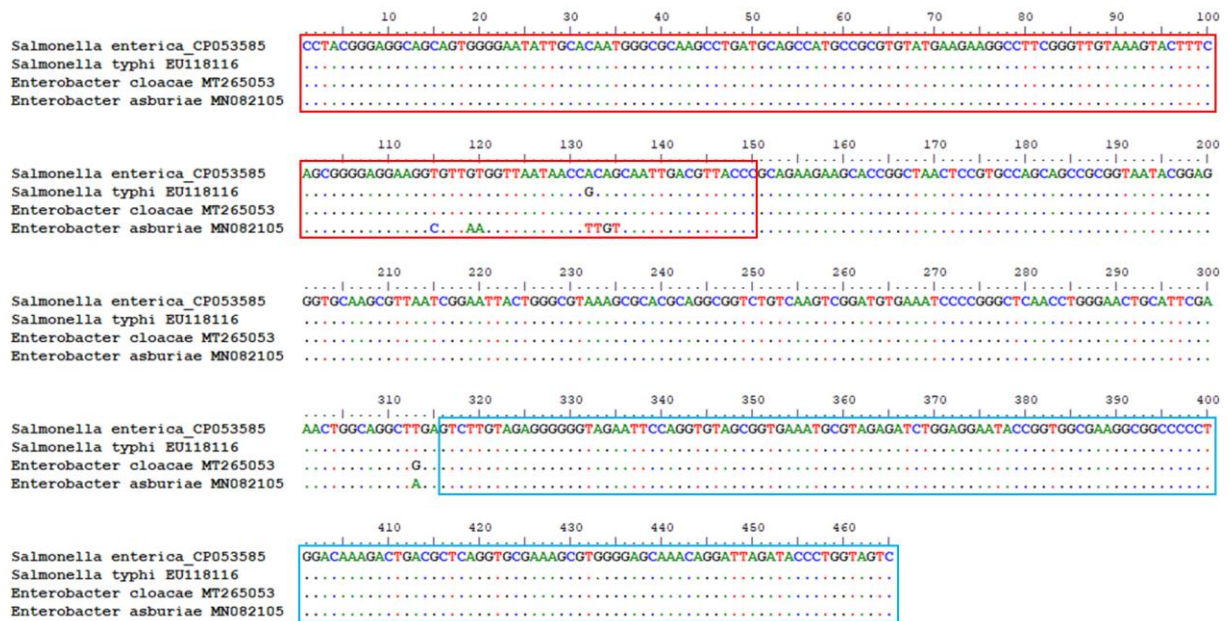


圖一、本計畫與Illumina公司對相同標準品ATCC MSA-1003 16S rRNA V3-V4區域總體基因體學之比較。

Input Pathogens*	Gram Stain	gDNA Abun. (%)	Characterization in this study	gDNA Abun. (%) in this study
<i>Pseudomonas aeruginosa</i>	-	12	<i>Pseudomonas</i>	10.4
<i>Escherichia coli</i>	-	12	<i>Escherichia</i>	5.1
<i>Salmonella enterica</i>	-	12	-	0
<i>Lactobacillus fermentum</i>	+	12	<i>Lactobacillus</i>	13.8
<i>Enterococcus faecalis</i>	+	12	<i>Enterococcus</i>	3.1
<i>Staphylococcus aureus</i>	+	12	<i>Staphylococcus</i>	14.8
<i>Listeria monocytogenes</i>	+	12	<i>Listeria</i>	14.6
<i>Bacillus subtilis</i>	+	12	<i>Bacillus</i>	19.1
<i>Saccharomyces cerevisiae</i>	Yeast	2	-	0
<i>Cryptococcus neoformans</i>	Yeast	2	-	0
			<i>Enterobacter</i>	4.4

\* : ZymoBIOMICS® Microbial Community Standard

表一、本計畫對測試品ZymoBIOMICS混樣核酸之16S rRNA V3-V4區域總體  
基因體學分析結果



圖二、部分 *Salmonella* 及 *Enterobacter* 菌屬細菌之 16S rRNA V3-V4 區域序列

比對結果

<b>Input bacteria strains</b>	<b>16S V3-V4 rRNA metagenomic</b>
<i>Corynebacterium striatum</i>	<i>Corynebacterium</i>
<i>Acinetobacter baumannii</i>	<i>Acinetobacter</i>
<i>Enterobacter asburiae</i>	<i>Enterobacter</i>
<i>Borrelia burgdorferi</i>	<i>Borrelia</i>
<i>Leptospira borgpetersenii</i>	<i>Leptospira</i>

表二、測試5種臨床細菌之16S rRNA V3-V4區域總體基因體學分析結果

ATCC MSA-1003	gDNA Abun. (%)	Characterization Species	Taxonomic classification(%) of 16S metagenomics
<i>Porphyromonas gingivalis</i>	18	<i>Porphyromonas gingivalis</i>	23.8
<i>Rhodobacter sphaeroides</i>	18	<i>Rhodobacter sphaeroides</i>	1.3
<i>Staphylococcus epidermidis</i>	18	<i>Staphylococcus epidermidis</i>	6.6
<i>Streptococcus mutans</i>	18	<i>Streptococcus mutans</i>	27.8
<i>Escherichia coli</i>	18	<i>Escherichia coli</i> <i>Escherichia albertii</i>	14.7
<i>Bacillus cereus</i>	1.8	<i>Bacillus cereus</i>	4.9
<i>Clostridium beijerinckii</i>	1.8	<i>Clostridium beijerinckii</i>	0.5
<i>Pseudomonas aeruginosa</i>	1.8	<i>Pseudomonas aeruginosa</i>	0.2
<i>Staphylococcus aureus</i>	1.8	<i>Staphylococcus aureus</i>	1.9
<i>Streptococcus agalactiae</i>	1.8	<i>Streptococcus agalactiae</i>	3.2

ZymoBIOMICS Input Pathogens	gDNA Abun. (%)	Characterization Species	gDNA Abun. (%) in this study
<i>Pseudomonas aeruginosa</i>	12	<i>Pseudomonas aeruginosa</i>	1.4
<i>Escherichia coli</i>	12	<i>Escherichia coli</i> <i>Escherichia albertii</i>	7.9
<i>Salmonella enterica</i>	12	<i>Salmonella enterica</i>	6.7
<i>Lactobacillus fermentum</i>	12	<i>Lactobacillus fermentum</i>	20.1
<i>Enterococcus faecalis</i>	12	<i>Enterococcus faecalis</i>	7.1
<i>Staphylococcus aureus</i>	12	<i>Staphylococcus aureus</i>	10.4
<i>Listeria monocytogenes</i>	12	<i>Listeria monocytogenes</i>	17.3
<i>Bacillus subtilis</i>	12	<i>Bacillus subtilis</i> species complex	14.5
<i>Saccharomyces cerevisiae</i>	2	-	0
<i>Cryptococcus neoformans</i>	2	-	0

表三、標準品(ATCC MSA-1003)與測試品(ZymiBIOMICS)之長片段16S

rRNA總體基因體學分析結果

## 109 年度科技研究計畫期末執行進度審查意見回復表

計畫名稱：高通量定序(NGS)平台開發與應用

計畫主持人：慕蓉蓉

填報日期：109.12.18

\*修正處請在報告中以紅字標示

序號	審查意見	辦理情形說明	修正處頁碼
1	無特殊建議。		
2	成 NGS 平台於不明原因感染原之檢驗及流程，進而完成全基因組裝定序，用以對 NDM-5 群聚之分析，進而建立 16S rRNA 總體基因體學技術，持續精進 NGS 檢測技術。	謝謝委員肯定。	
3			
4			

備註：如有修正期末報告內容，請註明頁碼，並務必於 109 年 12 月 23 日前至 GRB 系統完成資料抽換。