

計畫編號：MOHW107-CDC-C-315-124511

衛生福利部疾病管制署 107 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

107 年 度/全 程 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：107 年 1 月 1 日至 107 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		7
四、討論		10
五、結論與建議		11
六、參考資料		12
六、圖、表		14

共 (20) 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，下世代定序

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、 檢體來源：本署法傳系統通報之細菌株。
- 2、 檢體核酸萃取：利用自動核酸萃取系統 MagNA Pure Compact Instrument (Roche Applied Science)進行檢體核酸萃取，萃取完成的核酸置於-80°C 冷凍櫃保存。

本實驗方法以兩種平台進行序列分析。

- 3、 Illumina MiSeq 系統進行高通量定序，利用 QIAseq FX Single Cell RNA Library Kit 進行建庫，並以 Illumina MiSeq 平台進行 2x150 bp 定序。
- 4、 採取 PacBio RS 系統進行高通量定序，由於 PacBio RS 每段序列讀長至少 20 kb，較能得到完整組裝的全基因體全貌，每次實驗核酸需求量需 30 ug 或 4 ug，一次實驗之總輸出約 300 Mb 以上的核酸序列。
- 5、 由上述之兩平台結果互相校正，得出正確組裝數據。
- 6、 完成之片段再以 ResFinder 3.0 (Center for Genomic Epidemiology, DTU, Denmark 進行抗藥基因分析。

三、 結果

2011年德國豆芽菜遭大腸桿菌污染，造成多國嚴重群聚，事件發生後，世界各國以全基因定序分析此為新的菌株，為帶有出血性大腸桿菌(EHEC)毒素基因質體平行移轉(horizontal transfer)至附著型大腸桿菌EAEC，產生之新型菌株。本署負有傳染病防疫之使命，更需建置NGS平台，在面對新變異病原出現時，能及時進行全基因組裝解密，以及時提供防疫政策之實施。

本計畫107年以同時產生carbapenemase與MCR-1之多重抗藥細菌為研究對象，此種抗藥基因大多由質體攜帶，傳播快速，且容易產生變異型variant，一但傳播，將面臨無藥可用的窘境。因此選擇一株同時帶有NDM-9與MCR-1之大腸桿菌進行全基因解密。

1、 帶 MCR-1 與 NDM-9 基因的大腸桿菌基本資訊

(1). 藥敏測試：

本次選取同時帶 MCR-1 與 NDM-9 基因的大腸桿菌之藥敏結果如表一，該菌株對多種抗生素具有抗藥性，包括 carbapenem、colistin、cefepime 及 ciprofloxacin 等，僅對 amikacin 及 tigecycline 敏感。

(2). 質體分析：

以 S1-PFGE 將質體 linearization，再分別使用 MCR-1 及 NDM-9 探針進行 Southern hybridization，結果顯示 MCR-1 質體大小約 60kb，NDM-9 質體大小約 220kb，兩者位在不同的質體上。

(圖一)

2、以 NGS 平台建立全基因定序組裝

- (1). 以第二代 NGS 系統 Illumina MiSeq 平台進行全基因序列分析，本次共讀取約 232 萬條 reads，經 quality trimming 後得到近 186 萬條符合品質的 reads，經 *de novo* assembly 後共組裝出 156 條 contigs。(表二)

經 MLST 1.8 及 PlasmidFinder 1.3 分析，該菌株型別為 ST617，且包含 IncFIA、IncFII、IncHI2、IncHI2A、IncI1 及 IncI2 等 Plasmid 型別。ResFinder 3.0 分析，該菌株確實具有 NDM-9 及 MCR-1 基因，此外也包含 *bla*_{CTX-M-65}、*aac(3)-Iva* 及 *sul1* 等多種抗藥基因。(表二)

- (2). 由於 Illumina MiSeq 平台僅能讀取短片段 reads，對於長度較長且有多個重複片段（如 insertion sequence）的樣本，將大幅影響組裝效率，因此搭配第三代 PacBio 之 NGS 定序平台，提高

組裝成功率。PacBio 有長片段序列分析之優勢，但因其錯誤率高，所以組裝之 contig 需以第二代精準度高之 miseq 進行校正。

以 PacBio 組裝結果，得到多個 contig，經過與 miseq 序列校正後，其中兩個完整 contig 為細菌 chromosome，大約 5 Mb，以及 60 Kb 大小之帶 MCR-1 質體 (圖二)。另外，帶有 NDM-9 約 220 Kb 之質體，完成 154 Kb 之部分組裝。

3、全基因質體序列比對

- (1) 帶MCR-1質體基因序列與NCBI中6個序列有高度相似性，分別是來自中國大陸、南韓、馬來西亞以及澳洲之帶MCR-1質體基因序列(圖三)。這些質體除MCR-1外，沒帶有其他抗藥基因。
- (2) NDM-9 全基因資訊在NCBI資料庫中較少，無法找出相似度高之基因序列。帶NDM-9質體154 Kb基因序列中，有150Kb與一非帶NDM-9之質體相似(pHNYJC8)。而含NDM-9附近基因處有18Kb與一個 *Salmonella enterica* 帶有之NDM-9質體相似(pC629)(圖四)。

四、討論

- 1、 本計畫分析之大腸桿菌帶有 MCR-1 及 NDM-9 質體。此 MCR-1 質體與國際上其他國家釋出之 6 個質體相似，但與第一株發現之序列，缺少 IS*Apl1* insertion element (圖三標示)。IS*Apl1* insertion element 主要可使基因易位(translocation)，傳播至其他序列中。若未來因 recombination 獲得此易傳播 element，傳播至其他抗藥質體或 chromosome 上之可能性將大大提高，造成用藥上之困難。
- 2、 帶有 NDM-9 基因序列之資訊並無太多，但比較顯示，本計畫之 NDM-9 質體中，NDM-9 附近亦有 IS26 及 class 1 intergron 基因存在，顯示此 NDM-9 抗藥基因，可藉由形成 transposon 而傳播至其他質體上及細菌中，造成抗藥性嚴重問題。

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，第一 (106) 年已完成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。本(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。
- 2、 本年度建置之流程，在未來面對新型變異之病原，提供全基因序列解密，更可與國際序列進行比對，與國際接軌。
- 3、 本計畫第三(108)年，將利用 NGS 產出之大量序列，進行群聚事件追蹤分析。

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] Takeuchi F, Sekizuka T, Yamashita A, Ogasawara Y, Mizuta K, Kuroda M. MePIC, metagenomic pathogen identification for clinical specimens. *Japanese journal of infectious diseases*. 2014;67:62-5.

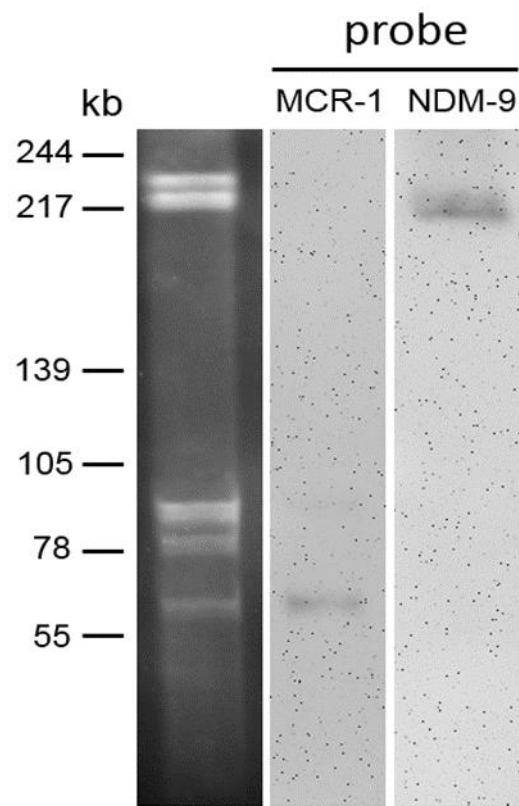
[14] Turner S, Pryer KM, Miao VP, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *The Journal of eukaryotic microbiology*. 1999;46:327-38.

[15] Zaravinos A, Bizakis J, Spandidos DA. Prevalence of human papilloma virus and human herpes virus types 1-7 in human nasal polyposis. *Journal of medical virology*. 2009;81:1613-9.

七、圖、表

Antimicrobials ^o	MIC(mg/L) ^o
<u>Cefepime</u> ^o	>16 ^o
Ceftazidime ^o	>32 ^o
<u>Aztreonam</u> ^o	>16 ^o
<u>Amoxicillin-Clavulanate</u> ^o	>16/8 ^o
<u>Ticarcillin-Clavulanate</u> ^o	>64/2 ^o
Piperacillin/ <u>Tazobactam</u> ^o	>64/4 ^o
Imipenem ^o	>8 ^o
<u>Meropenem</u> ^o	>16 ^o
<u>Ertapenem</u> ^o	>2 ^o
Ciprofloxacin ^o	>2 ^o
Chloramphenicol ^o	>32 ^o
Amikacin ^o	<=4 ^o
Gentamicin ^o	>8 ^o
Tetracycline ^o	>16 ^o
<u>Tigecycline</u> ^o	1 ^o
<u>Colistin</u> ^o	>4 ^o

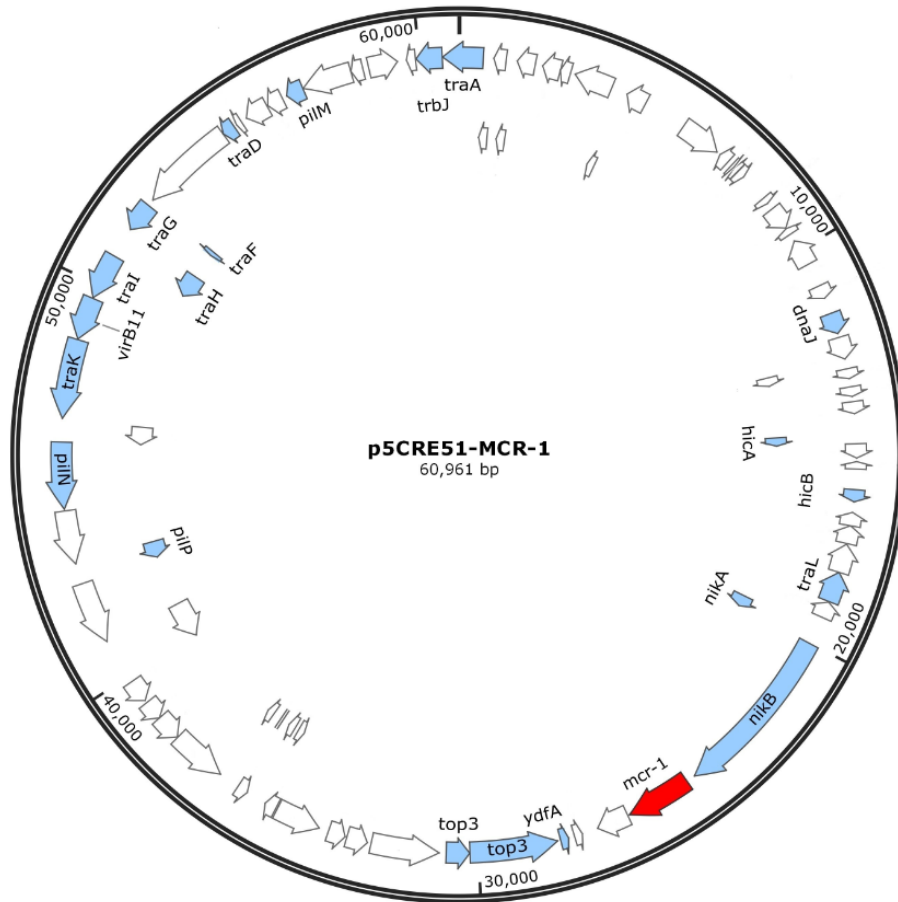
表一、同時帶有 NDM-9 及 MCR-1 基因之大腸桿菌敏結果



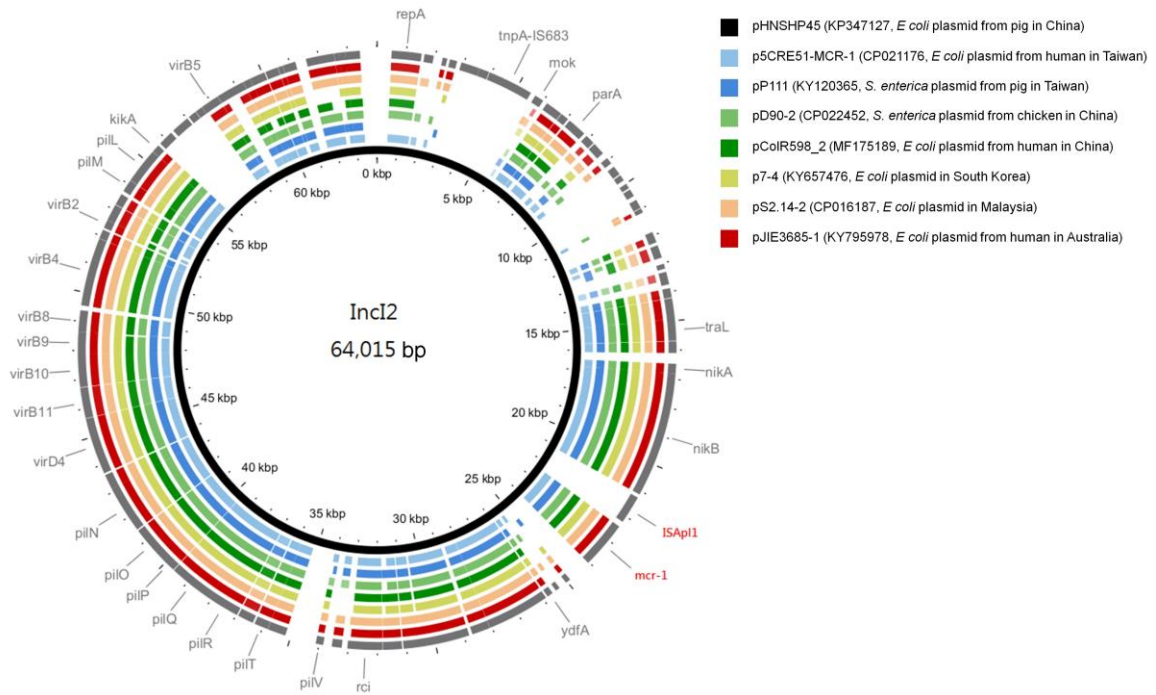
圖一、大腸桿菌質體分析

項目	分析結果
Total reads	2,324,564
Quality trimming	1,856,910
No. of <u>contigs</u>	156
MLST ST type	ST617
Plasmid <u>Inc</u> type	<u>IncFIA</u> , <u>IncFII</u> , <u>IncHI2</u> , <u>IncHI2A</u> , <u>IncI1</u> , <u>IncI2</u>
Antimicrobial resistance gene	<i>bla_{NDM-9}</i> , <i>mcr-1</i> , <i>aac(3)-IVa</i> , <i>aph(4)-Ia</i> , <i>bla_{CTX-M-65}</i> , <i>cmlA1</i> , <i>dfrA12</i> , <i>erm(B)</i> , <i>floR</i> , <i>fosA</i> , <i>strA</i> , <i>strB</i> , <i>sul1</i> , <i>sul2</i> , <i>sul3</i> , <i>tet(B)</i>

表二、Miseq 序列產出與基因分析

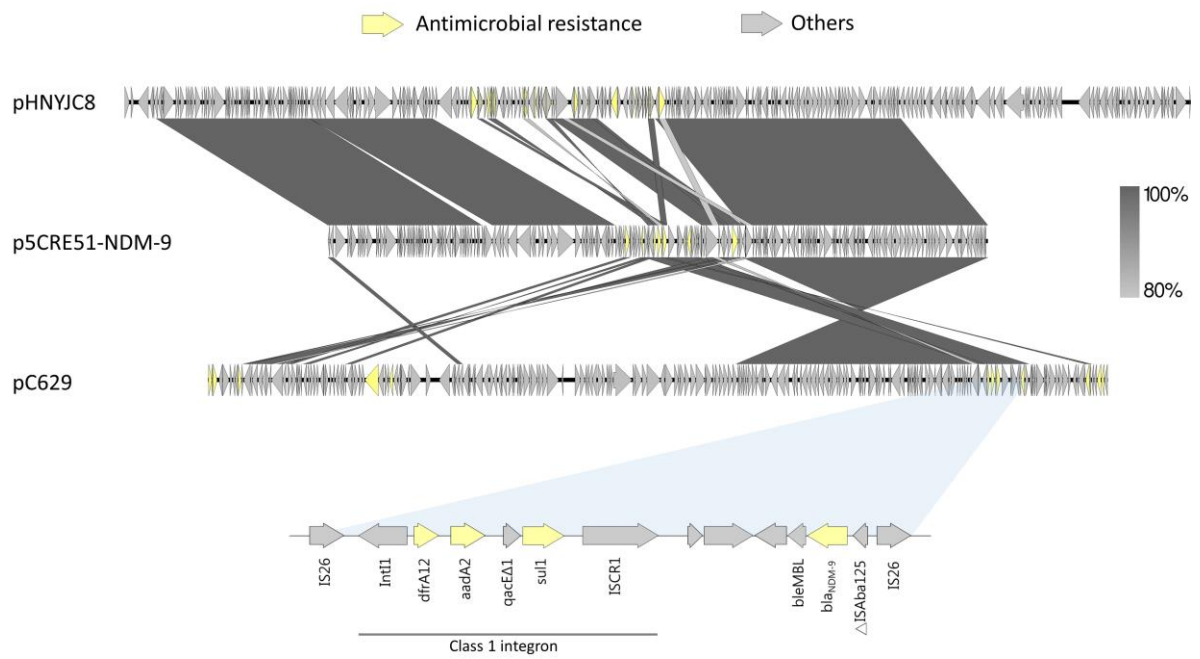


圖二、MCR-1 質體全基因定序組裝與基因分析



圖三、MCR-1 質體與國際間質體比對

圖二、未知感染源之檢驗流程圖



圖四、帶NDM-9質體之比較分析

衛生福利部疾病管制署 107 年科技研究計畫

期末審查意見回復

計畫編號：MOHW107-CDC-C-315-124511

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

*修正處在報告中加底線標示

序號	審查意見	主持人回復說明	修正處頁碼
1	以 NGS 平台組裝分析同時表現 MCR-1 與 NDM-9 之 <i>E. coli</i> ，釐清其傳播與抗藥性之特質。	謝謝委員	
2	可對高度抗藥菌株之防治政策制定有幫助。	謝謝委員	
3	達成原設定目標，未來有發生疫情，可透過全基因體定序組裝了解病原菌。	謝謝委員	
4	建議補充 NGS 應用於抗藥性監測或確立之優劣	謝謝委員建議，待資料分析後，將補充 NGS 應用於抗藥性監測或確立之優劣。	
5			