

計畫編號：MOHW106-CDC-C-315-113134

衛生福利部疾病管制署 106 年署內科技研究計畫期末成果報告

結核菌全基因體檢驗技術應用於結核病監測

研究報告

執行單位：衛生福利部疾病管制署研究檢驗及疫苗研製中心

計畫主持人：周如文

計畫協同主持人：莊珮君

研究人員：翁瑞芸

執行期間：106 年 01 月 01 日至 106 年 11 月 30 日

\*本研究報告僅供參考，不代表本署意見，如對外研究成果應事先徵求本署同意\*

## 目 錄

	頁碼
封面	1
目錄	2
壹、中英文摘要	5
貳、本文	
一、前言	9
二、材料與方法	14
三、結果	19
四、討論	26
五、結論與建議	30
六、計畫重要研究成果及具體建議	32
七、參考文獻	33
八、圖、表	
表一、本計畫評估之第二代與第三代全基因定序平台	37
表二、挑選 2009-2014 抗藥性聚集菌株	37
表三、TGS-TB 全基因定序抗藥性分析與傳統藥敏結果比較	38
表四、TB profiler 提供抗藥基因分析與疾管署分子抗藥檢測服務比較	39
表五、疾管署分子抗藥檢測分析與傳統藥敏結果比較	40
表六、RIF 抗藥位點統計	41

表七、INH 抗藥位點統計	41
表八、EMB 抗藥位點統計	41
表九、SM 抗藥位點統計	42
表十、二線針劑藥物 CAP、KM、AM 抗藥位點統計	42
表十一、FQ 抗藥位點統計	42
表十二、ETH 抗藥位點統計	43
表十三、PZA 抗藥位點統計	43
表十四、各聚集菌株地理區分佈與 spoligotyping 實驗結果	44
表十五、C00002 聚集菌株 TGS-TB IS6110 插入位置分析表	46
表十六、C00003 聚集菌株 TGS-TB IS6110 插入位置分析表	46
表十七、C00009 聚集菌株 TGS-TB IS6110 插入位置分析	47
表十八、C00015 聚集菌株 TGS-TB IS6110 插入位置分析	47
表十九、C00017 聚集菌株 TGS-TB IS6110 插入位置分析	48
表二十、C00024 聚集菌株 TGS-TB IS6110 插入位置分析	48
表二十一、C00037 聚集菌株 TGS-TB IS6110 插入位置分析	49
表二十二、C00041 聚集菌株 TGS-TB IS6110 插入位置分析	50
表二十三、C00042 聚集菌株 TGS-TB IS6110 插入位置分析	51
表二十四、C00048 聚集菌株 TGS-TB IS6110 插入位置分析	52
表二十五、C00049 聚集菌株 TGS-TB IS6110 插入位置分析	52
表二十六、C00051 聚集菌株 TGS-TB IS6110 插入位置分析	53
表二十七、C00065 聚集菌株 TGS-TB IS6110 插入位置分析	53
表二十八、C00070 聚集菌株 TGS-TB IS6110 插入位置分析	54
表二十九、C00074 聚集菌株 TGS-TB IS6110 插入位置分析	54
表三十、C00113 聚集菌株 TGS-TB IS6110 插入位置分析	55
表三十一、C00041 PacBio 分析挑選菌株抗藥圖譜	56

表三十二、PacBio 定序 C00041 挑選菌株之基因體組裝結果	56
表三十三、mCDC108 Illumina 定序結果比對 PacBio 定序深度小於 40%區域	56
表三十四、DNA 萃取方法比較	57
表三十五、結核菌 Illumina 全基因定序分析工具	57
圖一、C00042 中的混合菌株 spoligotyping 預測與實驗結果比較	58
圖二、C00002 聚集菌株 SNV 差異數及演化樹分析	58
圖三、C00009 聚集菌株 SNV 差異數及演化樹分析	58
圖四、C00015 聚集菌株 SNV 差異數及演化樹分析	58
圖五、C00041 聚集菌株 SNV 差異數及演化樹分析	59
圖六、mCDC108 PacBio 比對 Illumina 結果	59
圖七、C00041 聚集菌株抗藥位點與演化樹對照圖	59

## 壹、中英文摘要

本計畫目的為評估全基因體定序 (whole genome sequencing, WGS) 檢驗技術應用於結核菌之基因分析，並評估各主要定序平台之各項基因分析效益，以建立結核菌全基因體檢驗策略與流程，應用於臺灣結核菌抗藥性、基因型等基因特性監測，並有效輔助結核病聚集事件傳染源追蹤、空間分布、時序性，達到即時監測及降低結核病傳播的最終目標。結核菌全基因體約 440 萬個核酸，GC 百分比為 65.6%。自 2010 年起，國際學者開始利用全基因體定序技術試圖瞭解結核病傳播的問題，特別是在發生率較低的國家。當流行病學調查資料不盡完整時，利用全基因體定序技術與資訊分析，確實有極大的助益。近幾年由於全基因體定序相關平台與技術的蓬勃發展，費用已逐漸降低，並提高檢測的敏感度，因此大幅被許多國家所評估與使用。若整併入結核病個案管理資料系統，執行較大規模結核菌監測，早期預測或發現感染源，採取必要的防治介入措施。

**計畫目的：**全基因體定序可同時獲得抗藥性、基因型及單核苷酸多態性等核酸資訊。因此建立利用單一的全基因體定序檢測流程，可獲得有效的檢驗策略，除可提升診斷效能，並可縮短結核病人發現與治療及可能繼續傳播的時間，達到結核病發生率下降的目的。

**實驗方法：**選取基因型及不同抗藥型態代表性菌株，進行全基因體定序及分析，並制定各項判定準則，比較國際資料庫。選定已知群聚及高發生率地區菌株，執行全基因體定序研究，並執行感染源分析。

**研究結果：**建立完成 Illumina 全基因定序結核菌之 DNA 萃取、樣本庫置備、資料分析的流程。選取 56 株台灣盛行之多重抗藥性聚集菌株，評估線上結核菌全基因定序分析工具之效能。在抗藥性、基因型以及群聚傳染源分析上，可提供比現行疾管署檢驗項目更完整與詳細的資訊。另外在 PacBio 定序平台進行長讀長定序與基因體組裝，取得臺灣地區高盛行 MDR 聚集基因型之參考序列，可做為優化定序流程，與建立各基因型參考序列之基礎。

**結論與建議：**全基因體定序在結核病抗藥性檢驗與聚集判定疫情上，極具應用價值，然而在樣本製備流程及資料分析仍有待優化。依據本計畫第一年所得研究結果，建議(1)更改 Illumina 讀長及定序深度，以符合經濟效益與分析需求；(2)積極規劃全基因定序抗藥性與基因型之資料架構，以串連檢驗結果至傳染病通報與動態即時監測系統，並評估其它極具臨床檢驗應用性的全基因定序平台，縮短檢驗時效。

關鍵字：結核病、全基因體、監測

## **Abstract**

The goals of this project are to evaluate the feasibility of the whole genome sequencing (WGS) technique for tuberculosis (TB) control program in Taiwan; to assess the feasibility of **single nucleotide polymorphism**, indel, LSP analysis based on several current sequencing platforms for establishing the WGS-based strategy in surveillance of drug resistance and genotypes of *M. tuberculosis*; to identify the molecular markers to infer epidemiological links and outbreak events to improve the performance of epidemiologic investigations and TB control. The whole genome of *M. tuberculosis* contains approximately 4.4 million nucleotides, and the GC content is 65.6%. WGS analysis provides considerable benefit while epidemiological data are not available. Since 2010, researchers started using WGS techniques to resolve the transmission of TB and understand mechanisms of drug resistance, particularly in low incidence countries, including UK, USA, Canada, Germany and the Netherlands. Genetic information including drug resistance, genotypes, and nucleotide polymorphisms will be available simultaneously. We might be able to establish a rapid and effective drug resistance detection assay using WGS to increase the efficiency of TB management. Therefore, the duration of case finding will be shortened and patients will be treated promptly to reduce the possibility of transmission. We (1) select representative strains for sequenced and analyzed; (2) investigate genetic factors that enhance the chance of an outbreak; (3) understand possible transmission routes. WGS information will be incorporated into the TB case management systems (such as GIS, etc.). These can aid epidemiologists to study outbreak events and to predict or find the source of infection.

## **Methods**

In the first year, we select drug-resistant *Mycobacterium tuberculosis* isolates with representative genotypes in Taiwan. Using the WGS data to establish the analysis guidance and comparing to the result with global databases. Choose clustered isolates to construct the

infection chains.

### Results

We establish standard operation procedure DNA extraction, library preparation, data analysis using the Illumina platform. Of the 56 clustered multidrug-resistant *M. tuberculosis* isolates, the WGS analysis results provides more detailed information than current tests using in drug resistance, genotype, and phylogenic analysis. While the Pacbio platform with the long-read sequencing technology and de novo assembly, we obtain the reference genome of clustered MDR strains.

### Conclusions and suggestions

The WGS is valuable in TB diagnosis and outbreak investigations, however the sample preparation procedure and data analysis are still need to be improved. According to the results, we suggested that the Illumina sequencing read length and coverage depth need to be modified to fit the economic considerations and analytic requirements. And we also recommended to evaluating the new and simple WGS platforms with better applicable potential in clinical diagnosis.

**Keywords:** tuberculosis, whole genome sequencing, surveillance, drug resistance



## 貳、本文

### 一、前言

生醫發展的關鍵六大技術：次世代定序、蛋白質體學、liquid biopsy、CRISPR 基因修飾技術、3D bioprinting 及大數據分析，結合資訊及各類疾病工具之嶄新健康促進模式，可望讓醫療進入「即時醫學」和「預防醫學」的時代。其中，運用基因體定序平台技術整合生物資訊分析，使得各類遺傳性疾病、慢性疾病、或感染性疾病，都可進行即時診斷、療效評估及公衛管理，可達到有效疾病防治及保健的願景。

為有效降低結核病的發生率之方法主要可由三方面介入：(1)即早發現病人、(2)適當治療及(3)阻斷結核病傳播。世界衛生組織估計在全球所有約 900 萬結核病人中，仍有三分之一的病人未被診斷發現或未被通報；而在 2013 年預估 48 萬的多重抗藥性 (multidrug-resistant, MDR) 結核病人中，也僅約 45% MDR 病人被診斷出來【1】。隨著全基因體定序 (whole genome sequencing, WGS) 技術在結核菌的分析應用成功後，國際間紛紛利用 WGS 快速鑑別與抗藥性相關的突變位點【2-5】；另外也利用 WGS 彌補目前以 MIRU 及 spoligotyping 為主要方法進行基因型分析，判定結核病傳播時所遭遇的盲點【6-12】。

Foundation for Innovative New Diagnostics (FIND) 已明確的揭示將在 2018 年，建立抗藥性基因序列分析的黃金標準 (gold standard)，期望能以分子基因序列分析結果，快速診斷結核菌抗藥性，以縮短需等待傳統費時的藥物感受性試驗 (drug susceptibility testing, DST)。緣此，則必須藉由進行大規模的全基因體定序分析 (預估超過全球各區 5 萬筆全基因體資料)，以正確分析各抗藥性相關基因突變位點資料與臨床治療結果的相關性，進而彌補分子診斷檢測與傳統 DST 結果之間的誤差，以建立黃金標準。雖然世界衛生組織自 1994 年，即開始進行全球結核病抗藥性監測，但因受限於傳統 DST 方法學的限制，在抗藥性監測資料仍非全面性，因此世界衛生組織也在近年開始導入，以抗藥性相關基因分析進行抗藥性監測。首先是 pyrazinamide (PZA) 及 fluoroquinolone (FQ) 類藥物之抗藥性監測，選擇 5 個國家進行全國性或區域性的分析，由區域參考實驗室 (supranational reference laboratory, SRL) 進行傳統 DST 及對應抗藥性基因 pncA 及 gyrA 及 gyrB 序列分析。將全基因體序列分析結果與傳統 DST 進行比對，

PZA 藥物的分子抗藥性偵測 sensitivity 為 80-90%及 specificity 為 99%【13】。此監測研究持續進行中，也將再擴及其他地區。並利用其他抗藥相關的基因序列分析，如：rpoB、inhA 及 katG 等進行監測。利用抗藥性基因分析，除可進行即時性監測，也可改善建立及標準化新藥物傳統 DST 方法學所需的時間。此外，臨床治療結果資料的收集，仍須與基因資料比對，比對結果才能正確確認抗藥基因與實際抗藥的關聯性，也才能實際利用分子快速診斷方法，建立未來分子抗藥性檢測基準。目前利用全基因體定序分析技術進行的抗藥性監測，仍侷限於已知的抗藥性相關基因，如：rpoB、katG、inhA promoter、pncA/promoter、gyrA、gyrB 及 rrs 等，尚未有新的基因位點被推薦使用；主要是因為藥物作用機制與抗藥性相關基因的關聯，尚未能全盤了解。只有持續累積大規模的全基因體定序分析資料時，才有助於釐清問題。例如，研究利用全基因體序列分析得知，在 rifampicin (RIF)抗藥相關的 rpoB 熱點突變也可藉由伴隨 rpoA、rpoC 補償性基因突變，增加抗藥性菌株的適應性 (fitness) 進而增加 MDR 菌株的傳播【14】。

由於全基因體以次世代定序技術 (next generation sequencing, NGS) 之序列分析，敏感度遠較傳統定序的 Sanger sequencing 方法高，因此利用全基因體分析，也有較高敏感度來偵測 hetero-resistant 的存在。在傳統 DST 的檢驗中，有可能因為抗藥與敏感菌存在的比例量不同而被選殖，可能導致錯失檢測抗藥菌。僅有累計大量代表性及多樣性的檢體進行全基因體序列分析，並與治療結果比對，才可找出新的抗藥相關基因突變位點，納入抗藥性檢測標的，以確保提高 sensitivity、specificity 及陽性預測值 (positive predict value)，如此將可確實縮短抗藥性病人診斷時間，給予適當治療，以減少抗藥性病人的傳播機會。

自 2010 年起，國際間開始利用全基因體定序技術，進行結核病傳播與聚集事件的調查研究【6-12】。因基因序列的 single nucleotide polymorphism (SNP) 改變是具有時序性，除可獲得不同菌株間的基因 SNP 差異外，也可獲知疾病傳播的空間方向性與時序性，進而追蹤到指標個案(index case)或者是超級傳播者(super spreader)。因此，對結核病的傳播與防治可提供相當完整的資訊。在結核病低發生率的國家如英國，已開始針對每一個結核病人的結核菌株進行全基因定序分析。在低發生率的國家因為接觸者追蹤，相較於高發生率的國家困難，因此全基因體的分析是利器。利用全基因體定序分析資料也可幫助掌握結核病在該國境內的傳播途徑，並可釐清境外移民對該國結核病發生率的影響，進而有不同的防治策略與作為。然而低發生率與高發生率國家傳播的結核菌其 SNP

變化的速率可能不同，也可能與宿主及環境有關。因此到目前為止，儘管各國皆有以全基因體序列分析進行群聚事件的評估，也發現確實比現行的基因分型如 MIRU 的分析結果正確，可排除非真正的聚集事件，然對於菌株間 SNP 差異多少以上視為相同或不同基因型仍尚未有定論【6-12】。

由於全基因體定序分析技術於近年的快速進展，也因為所需時間與費用已大幅下降【15】，包括對分子快速檢測、藥物開發、抗藥性監測、菌株基因型監測、聚集事件調查等勢必有很大的助益。由於全基因體序列分析所需時間與花費的大幅減少，相對可進行大規模的檢體基因分析，並進行 phenotype 及臨床治療結果與 genotype 的比對，以更加確定分子抗藥性快速診斷的正確性與重要性，也才能達到世界衛生組織全球即時抗藥性監測的目的。菌株的全基因體序列分析可提供菌株所有基因的全貌，相較於現行的 MIRU 基因分型技術進行聚集事件分析，仍需配合疫調資料才可進行最後判定。因此在疫調難以完整進行或是需耗費大量人力與時間執行的情況下，全基因體定序分析可直接呈現所有的基因資料，並同時提供抗藥性、免疫性、毒性或其他病原體特性的資訊，可加速結核病的積極防治、診斷試劑開發與藥物的發展。因此藉由本計畫的執行，期能擬訂將全基因體定序分析檢驗技術納入實驗室檢驗或研究的策略，並評估建立抗藥性與聚集事件監測的可行性，以有效協助結核病的防治。

在定序平台的選擇上，基因體定序技術由第一代樣本庫載體為細菌質體近行多菌株傳統 Sanger 定序，演進到次世代定序 NGS 以橋式聚合酶鏈鎖反應(PCR)等技術在晶片平面放大單一樣本庫 DNA 進行高通量定序，近 5 年來第三代定序技術發展也亦趨成熟，如 Pacific Biosciences (PacBio) 使用的 SMRT (single molecule, real-time)單分子即時定序技術，以及不使用聚合酶鏈鎖反應、直接檢測 DNA 分子通過奈米孔各鹼基電位的 Oxford Nanopore 定序技術，提供研究者更合乎需求的技術選擇【16】。本研究選用市佔率最高、應用最廣泛的第二代定序平台 Illumina，與第三代定序平台 PacBio 進行結核病全基因體定序(表一)。

Illumina 平台原理為將全基因體 gDNA 打斷為短片段(300~500 bp)，並於片段頭尾接上 adapter 序列，再將此片段化的 DNA 樣品載入表面帶有互補 adapter 序列的晶片上，經過橋式聚合酶鏈鎖反應擴增待測 DNA 後，以標定有不同且可移除螢光基團的 dNTP，進行重複螢光標記、移除與偵測的 DNA 聚合反應，最後得到坐落晶片不同區域 DNA 所

測得的定序序列總合。其優點為 DNA 樣本需求量較低、適合快速對參考序列進行比對 (mapping) 分析。然而對於高 GC 含量 DNA 序列定序效率不佳，短定序讀長無法跨越高重複片段以及基因重組區域，基因體組裝 (*de Novo assembly*) 效率較差。

PacBio 原理為使用直徑 70 nm，深度 100 nm 微小孔洞所組成的光波導式感測晶片 (SMRT cell)，每一孔洞底部固定有一 DNA 聚合酶，當進行聚合反應時，不同的 dNTP 會釋出其標定有螢光基團的磷酸根，再經由僅能穿透孔洞底部約 20-30 nm 的激發光所激發，最後以感光元件偵測並紀錄螢光訊號的及時影像。上機樣本庫 (library) 的製備上，待測 DNA 片段兩端接上 hairpin adapter，定序反應時雙股 DNA 會先解旋為單股的環狀 DNA，再與 DNA 聚合酶反應。由於 SMRT 定序過程產生的隨機誤差率約 15%，利用重複偵測此環狀 DNA 的序列訊號，增加此 Circular consensus sequencing (CCS) 的重複性以提高定序深度，可整合消除誤差率，使定序精準度達到 99.9%。SMRT 系統因直接對原始待測 DNA 進行定序，可避免 NGS 系統製備 DNA 樣本庫時，使用聚合酶連鎖反應擴增可能造成的序列錯誤，同時對於高 GC 含量的 DNA 序列也能維持良好的定序效率。PacBio 系統的定序讀長一般為 10-15 kb，依不同定序儀性能最高可得 70 kb 以上的讀長，適合對具有大量重複序列及 DNA 重組的物種進行基因體組裝，除了分析短片段的序列插入、缺失、突變外，更利於分析宏觀的長片段基因結構的組成與變異。其優點為適合分析帶有大量重複序列、高 GC 含量的結核菌群。對於具有不同重複與插入序列組成的各 lineage 菌株，較易從定序結果組裝回完整的基因體，可補足由於 illumina 技術限制而無法被分析的序列，利於結核病的傳染時序分析。而其缺點為 DNA 樣本的品質與需求量皆高。

本研究總目標為利用已發展成熟的全基因體定序，進行結核菌全基因定序分析，以達到：(1) 快速提供結核菌抗藥性結果，運用在多重及超級抗藥結核病人的臨床用藥上，以提高治癒率；(2) 以更完整及鑑別力高的結核菌基因訊息，引導公共衛生的調查，並可藉由基因的突變獲得時序性結果，了解對結核病傳播的先後順序，有效確認感染源及控制疾病的傳播；(3) 發現新基因資訊，提供基礎研究機構及生技製藥產業，從事抗藥機制研究及診斷試劑或新藥物開發。分年度階段性目標如下：

#### **106 年度：**

**目的：**評估不同全基因體定序平台，如 illumina Hiseq 或 PacBio 系統，制定最佳化分析流程與策略。

**工作項目：**

- 一、選取抗藥性與非抗藥性及不同基因型代表性菌株，進行 WGS 分析，建立全基因體定序分析方法及生物資訊分析系統。
- 二、進行抗藥性相關基因與基因型分析及驗證。

**107 年度：**

**目的：**利用第一年建立的最佳化全基因體定序策略，挑選(1)已知或進行中群聚事件菌株、(2)高發生率地區菌株、及(3)執行 WGS 研究，分析感染源。

**108 年度：**

**目的：**評估擴大執行全基因體定序之可行性與量能，並將基因資料整併入結核病個案管理系統，整合並建立基因大數據。

## 二、材料與方法

### (一) 材料

由 ATCC 等機構購置之標準分枝桿菌株及醫療院所結核病實驗室分送至疾病管制署進行複驗及寄存之結核菌株(含 MDR 及 XDR 等抗藥性菌株)。

### (二) 研究及實驗方法

#### 1. 菌株收集及分離:

將菌株次培養於 7H10 培養基或 Löwenstein-Jensen 培養基，進行增菌與純化。

#### 2. 基因分型

##### 2.1 結核菌群間隔寡核酸分子分型法(spoligotyping):

依據 Kamerbeek 方法，首先將菌株以商業化套組 (Ocimum Biosolutions) 進行聚合酶連鎖反應以放大基因組中之直接重覆 (direct repeat) 片段，放大後之螢光產物與預先標定有 43 個 spacer 的探針之尼龍膜進行雜交 (hybridization)，最後使用 X 光片偵測 spacer 之有無。Spoligotype 分類依國際準。基因型為 Unknown 者，乃國際 SITVIT 資料庫尚未定義的型別，而 undefined 者則為該 spoligotype pattern 未見於 SITVIT 資料庫中。

##### 2.2 MIRU-VNTR 基因分型

根據 2001 年 Supply 等人發表 MIRU-VNTR 分析位點進行分析，包含 15 個位點：MIRU 4、MIRU 26、MIRU 40、MIRU 10、MIRU 16、MIRU 31、MIRU 2、MIRU 23、MIRU 39、MIRU 20、MIRU 24、MIRU 27、ETR A、ETR B、ETR C。Multiplex PCR 引子分成 5 組，每一組 PCR 引子端分別以不同螢光染劑 (FAM、HEX、TEMRA) 標示，每一 PCR 反應包含 50 to 200 ng 的 DNA、0.5 U Taq (AmpliTaq Gold, PE Applied Biosystems)、1X PCR buffer、2.5 mM MgCl<sub>2</sub>、0.2 mM dNTPs (each) 與 0.4 μM primers (each)，PCR 反應條件為：95°C，10 min，1 cycle；95°C，1 min，59°C，1 min，

72°C, 1 min 30 sec, 35 cycles; 72 °C, 8 min, 1 cycle。將 PCR 產物加入 40 µl H<sub>2</sub>O 進行 5 倍稀釋，稀釋完之產物取 2 µl 與 8 µl ET-RAX 900 DNA size ladder 混合以 GE-Amersham MegaBACE 500 genetic analyzer 用 2% Metaphor 瓊膠進行分析，較長 PCR 產物之 MIRU-VNTR 位點以 Applied Biosystems 3500XL 進行分析。並根據 Supply 提供之 PCR 產物長度與相對應 repeat 數進行換算。

### 2.3 數學演算分析

Spoligotyping 及 MIRU-VNTR 基因型的實驗資料以 BioNumerics® 進行輸入及分析。

## 3. 全基因定序

### 3.1 Illumina 定序平台

#### I. DNA 萃取及樣本庫製備上機

(1) 傳統 phenol/chloroform 萃取 DNA: 經 80°C 去活化一小時之結核菌菌體，以 lysozyme 37°C 處理 2 小時破菌後，加入 proteinase K 與 SDS 混合，於 56°C 作用隔夜，分解 DNase。於破菌液中加入 CTAB/NaCl 混合，沉澱蛋白質與細胞壁多醣，以 chloroform/isoamyl alcohol (24 : 1)、phenol/chloroform/isoamyl alcohol (25 : 24 : 1)、chloroform 各萃取一次 DNA 至水相後，以 isopropanol 於 -20°C 沉澱 DNA 隔夜。DNA 沉澱以 70% ethanol 清洗兩次後陰乾，再以 1X TE 緩衝溶液回溶。

(2) 樣本庫製備試劑套組: TruSeq Paired-End DNA, Gel free。

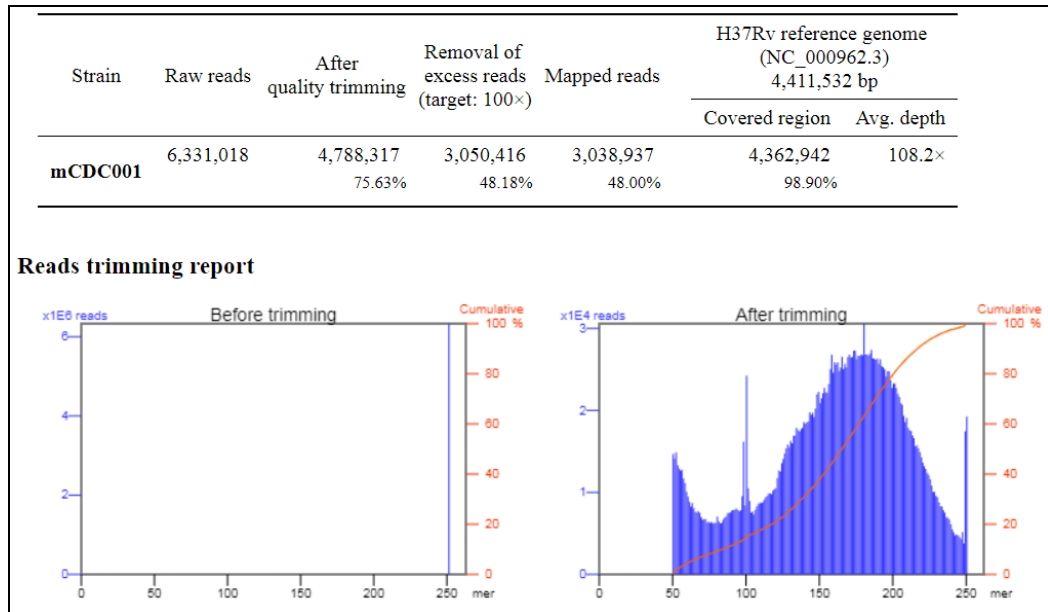
(3) Illumina HiSeq 2500 之定序模式: HiSeq Rapid v2, Paired End 2×250 bp，定序深度 240×。

#### II. Illumina 定序資料分析:

定序原始檔(reads data, 檔案格式 Fastq)，以 quality trimming 去除 adapter、定序品質不良、過短等序列後，與 *M. tuberculosis* H37Rv 參考序列比對(mapping)後可得到經排序後的 reads 檔案(檔案格式如.bam)，排序後的 reads 資訊經由演算可擷取出與參考序列相異的 SNPs、indel 清單(檔案格式: variant call format,

vcf)。

本研究使用日本感染症研究所開發的 Total Genotyping Solution for Mycobacterium tuberculosis, TGS-TB【17】網頁結核菌全基因分析工具套組，分析結核菌 Illumina 的定序品質、對參考序列的基因覆蓋率、平均定序深度，以及 10 種藥物的分子抗藥分析、基因分型預測。分析工具示意如下。



(1) TGS-TB quality trimming 與標準序列比對報告:

提供 quality trimming 前後、對參考序列 H37Rv 比對後(mapped reads)的序列數，比對覆蓋率(covered region)，及平均定序深度(avg. depth)。參考序列 H37Rv 比對僅使用 trimming 後 50-250 bp 的序列(Reads trimming report，右圖)，隨機選取至比對標準序列平均深度達 100x。



## SNV detection report

SNV sites	Number of SNVs		
	Original	Effective	
Original dataset	21,805	21,316	97.76%
Additional	-	152	
Total		21,468	

### Completed *in silico* genotyping

- [Potential AMR prediction](#)
- [KvarQ MTBC analysis](#)  
(Lineage / Antimicrobial Resistance)  
version: [MTBC.niid.3.20150403](#)
- [TBprofiler](#)
- [Core-genome phylogeny](#)
  - ML Tree by RAxML with bootstrap test 100 times
    - in [Newick](#) format (6.9KB)
    - in [PDF](#) format (8.2KB)
  - Linkage Network analysis (for [PopART](#))
    - [Queries with 78 references in NEXUS format](#) (1.7MB)
  - SNVs raw data
    - in [tabular](#) format (661KB)
    - in [multi fasta](#) format (1.7MB)
    - in [pairwise SNV matrix](#) (35KB)
- [Spoligotyping](#)
- [IS6110 insertion site](#)
- [VNTR \(our customised 43 loci\)](#) Download loci list in [PDF](#) or [Excel](#) format.

## (2) TGS-TB 抗藥性、演化樹與基因分型工具:

抗藥性分析工具包括了 Potential AMR prediction、KvarQ MTBC analysis，並提供英國開發之 TB profiler 分析結果，匯整抗藥相關基因與 lineage 相關 SNP 位點的偵測。AMR 抗藥分析之抗藥位點經日本感染症研究所合作之實驗室進行藥敏實驗，驗證其抗藥相關性。

在多菌株演化樹分析上，TGS-TB 選用了去除高變異性基因的 core-genome 區域，分析 21,805 個單核苷酸變異位點 Single Nucleotide Variant (SNV)，以及依據所上傳分析的多筆菌株資料，分析所得可提供演化鑑別的額外 SNV，進行傳播鏈與演化樹的分析運算。高變異性基因包含突變與基因重組頻繁區域，如與結核菌引發宿主免疫反應相關之 PE-PGRS 基因，以及抗藥性相關基因，或高重覆序列等。

基因分型預測則包含了 Spoligotyping、IS6110 插入點分析及 VNTR。

### 3.2 PacBio 定序平台

PacBio 上機樣品選用 Genra DNA purification (QIAGEN)試劑套組萃取基因體 DNA，萃取流程依廠商操作手冊進行。樣本庫依定序公司選用之上機模式製備，進行 SMRT 單分子及時系統定序，定序原始資料經 Hierarchical Genome Assembly Process (HGAP)演算及 circulator 校正進行 *de novo* 全基因體組裝。組裝後之基因體序列再以 WGS 資料分析軟體進行後續比對分析。

### 3.3 WGS 定序資料分析:

以 Geneious (Biomatters Ltd.)軟體進行定序序列品管(trimming)、參考序列比對(reference sequence mapping)、基因分析與註記(annotation)、基因變異(variant calling)等分析。分析檢測敏感性(sensitivity)、特異性(specificity)、陽性預測值(positive predictive value, PPV)及陰性預測值(negative predictive value, NPV)。

### 三、研究結果

#### (一) Illumina 定序平台效能評估

##### 1. 菌株挑選：

自 2009-2014 年已完成 IS6110 RFLP 或 MIRU(10)與 spoligotyping 合併基因型分析判定歸屬相同聚集(Cluster)的菌株中，選取不同基因型 16 個聚集之抗藥性結核菌群(*Mycobacterium tuberculosis* complex, MTBC)菌株共 56 株：包含 6 株超級抗藥 (extensively drug-resistance, XDR)及 50 株多重抗藥(multidrug-resistance, MDR)菌株。其中主要為北京型 24 株(42.9%)、Haarlem 型 16 株(28.6%)、EAI2-Manila 型 4 株及其他基因型(表二)。

##### 2. WGS 資料分析：

以 TGS-TB 線上基因體分析工具比對 *M. tuberculosis* H37Rv 參考序列，覆蓋率最低為 98.90%，最高者為 99.91%，且皆可達 TGS-TB 設定最佳的 100×平均定序深度。

##### 2.1 抗藥分析

利用 TB profiler 工具，由已知抗藥性相關位點分析 10 種 TB 治療藥物之分子抗藥性，並與傳統藥物敏感試驗結果(drug susceptibility testing, DST)比較。檢測敏感性(sensitivity) RIF、INH、EMB、KM、AM、FQ 皆為 100%，SM 為 79.2%、CAP 為 83.3%、ETH 為 85.2%、PZA 為 81.3%；特異性(specificity) AM、FQ、PZA 皆為 100%、EMB 為 83.9%、SM 為 96.9%、CAP 為 96.0%、KM 為 93.9%、ETH 為 55.2%；陽性預測值(PPV) RIF、INH、AM、FQ、PZA 皆為 100%，EMB 為 83.3%、SM 為 95.0%、CAP 為 95.0%、KM 為 70.0%、ETH 為 63.9%；陰性預測值(NPV) EMB、KM、AM、FQ 皆為 100%，SM 為 86.1%、CAP 為 98.0%、ETH 為 80.0%、PZA 為 93.0% (表三)。

比較現有疾管署所提供之分子抗藥檢測基因項目(表四):RIF 的 *rpoB* 檢測僅含蓋基因熱區 codon 426-452 (*E. coli* codon 507 至 533),V160F 位點非常規檢驗項目。使得 56 株菌株的檢出預測中,RIF 抗藥敏感性為 98.2%;INH 僅檢測 *inhA* (*fabG1* promoter)與 *katG* codon 315 前後序列,INH 抗藥敏感性預測值為 96.4%。而在 PZA 抗藥檢測中,由於 *pncA* 無明顯基因熱區,當啟動子與基因轉譯區發生突變,且胺基酸產生改變,視為與抗藥相關;與此同時,因 TB profiler 建立的 *pncA* 抗藥位點不足以含括所有可能抗藥突變形式,使得核苷酸 292-293 間的序列插入及 codon L120R 未被定義為與抗藥相關,疾管署的 PZA 分子抗藥檢測敏感性為 100%,而 TB profiler 僅為 81.3%(表五)。

56 株 MDR 挑選菌的抗藥位點統計中,RIF 以 *rpoB* S450L 41 株佔最多數(73.2%),基因熱區外 V170F 有 1 株(表六)。INH 抗藥位點中 *fabG1* c-15t 為最多數,共有 34 株(60.7%),其次為 *katG* S315T,共 28 株(50%),*fabG1* c-15t 合併 *katG* S315T 突變的菌株共 10 株(17.9%)、合併 *katG* A106V 者 1 株(1.8%),*fabG1* c-15t 合併 *inhA* I194T 突變位點則為 7 株(12.5%),其中有 4 株帶有 *kasA* G312S (7.1%),皆為 EAI2-Manila 世系。其餘有 2 株帶有 *fabG1* t-8c 突變 (3.6%),兩株 *katG* D735A (3.6%)(表七)。EMB 抗藥菌株中,帶有 *embB* D1024N 位點佔最多數共 10 株,其中合併 *embB* Q497P 位點者 9 株,再次為 *embB* M306V 共 8 株,帶有抗藥相關位點但傳統藥敏試驗結果為敏感共 5 株,分別帶有 *embB* M306V、M306L、G406D、Q497K、以及 *embA* c-12t 合併 *embB* G406S 突變(表八)。SM 抗藥菌株中,*rpsL* K43R 佔多數共 13 株,其中 6 株帶有 *rrs* a1401g 位點,此位點實為與二線針劑藥物抗藥相關,而與 SM 抗藥無關;其次為帶有 *rrs* a906g 者共 6 株(表九);SM 抗藥但未有 TB profiler 定義抗藥位點菌株的 5 株菌株,在更進一步分析其 *gid* 基因後,其中 1 株帶有 *gid* 基因在胺基酸 codon E170 突變為中止密碼子,並在該位點下游核酸序列 555 到 556 間發生單核苷酸 T 插入;1 株發生核苷酸 350G 刪除;1 株則有 50%的定序訊號帶有核苷酸 102G 刪除;另外有 2 株帶有 *gid* V77G 突變,

然而此突變在 C00041 聚集中的 11 株菌皆有發生，無法斷定與抗藥之相關性；*gid* E92D 突變為北京株之多型性基因位點，與 SM 抗藥無關【18】。二線針劑藥物的部分，CAP 抗藥菌株中，有 5 株帶有 *rrs* a1401g 位點，1 株在 *rrs*、*tlyA* 中無抗藥相關突變；2 株帶有 *rrs* a1401g 位點但傳統藥敏結果為敏感。KM 抗藥菌株皆帶有 *rrs* a1401g 菌株，而在 3 株非抗藥菌株中帶有 *eis* g-10c 突變，此位點尚未有文獻做深入探討，以目前結果推測與 KM 抗藥關連性低。AM 抗藥的 7 菌株皆帶有 *rrs* a1401g(表十)。FQ 抗藥菌株中，皆發生 *gyrA* D94 胺基酸突變，而又以 D94G 15 株為最多(表十一)。ETH 抗藥菌株中，23 株帶有 *fabG1* c-15t 突變，其中 5 株合併伴隨 *inhA* I194T 突變，而在 4 株未有 *fabG1*、*inhA* 突變之菌株中，3 株帶有 *ethA* 序列突變，1 株則未在 *ethA*、*ethR* 中發現可能抗藥相關之突變(表十二)。PZA 抗藥菌株則皆帶有 *pncA* 突變位點，以 *pncA* C14R 4 株所佔比例最高(表十三)。

本研究 TB profiler 所建立的抗藥位點 *kasA* G312S 被定義為 INH 抗藥相關，然有文獻證實該位點實為 EAI lineage marker【19】，與抗藥無關；另外，*rrs* a1401g 被定義為與 SM 抗藥相關，然而在先前的研究此位點僅被驗證與 CAP、KM、AM 抗藥相關【20】，將此位點修正為與 SM 抗藥無關後，TB profiler 的 SM 檢測特异性可由 96.9% 提升至 100%，陽性預測值由 95.0% 提升至 100%。

## 2.2 基因型分析

(1) spoligotyping 實驗結果與 WGS 預測結果比對:

除 C00042 聚集中一人帶有非 EAI2-Manila 基因型的混合菌株，WGS 分析結果與 spoligotype 結果不同外，其餘 55 人的分離菌株皆與 spoligotype 結果一致(表十四)(圖一)。

(2) IS6110 插入點分析同聚集個案菌株:

表十五至三十為各聚集菌株之 IS6110 插入點、插入方向及該插入點定序分析之覆蓋深度(count)一覽表，對照各菌株的全基因體平均定序深度(皆約 108 倍)，比值低表示該插入點前後可分析序列覆蓋率低，為訊號微弱區域，解析效力較差(比值低於 10%以粗體字表示)。

C00002 聚集中，3 菌株插入位置與方向一致，共偵測到 11 個插入點。

C00009 聚集中，5 菌株中有 1 菌株缺少 1 插入點訊號(H37Rv nt 3,123,045)，其餘插入位置與方向一致，共 12 個插入點。

C00003 聚集中，2 菌株插入位置與方向一致，共偵測到 11 個插入點。

C00015 聚集中，4 菌株插入點與方向一致，共 20 個插入點，其中有 3 菌株在 nt 2,165,962 的插入點覆蓋率低於 10% DP。

C00017 聚集中，共偵測到 20 個插入點，插入方向皆一致，其中 1 個插入點(nt 2,165,962)有 3 菌株覆蓋率低於 10% DP；另外有 4 個插入點在不同的菌株中缺少訊號如下，nt 2,639,539 有 2 菌株不含此插入訊號，其餘 4 株該位點覆蓋率低於 10% DP；nt 3,549,197 有 1 株帶有此插入訊號，其餘 5 株則在相隔 3 bp 的 nt 3,549,200 有插入訊號；nt 4,3319,652 僅 1 株帶有此插入訊號，並且覆蓋率低於 10% DP。

C00024 聚集中，nt 2,165,962 兩菌株中僅 1 株菌帶有此插入訊號，另 1 株菌該插入點覆蓋率低於 10% DP，其餘插入位置與方向一致，共 19 個插入點。

C00037 聚集中，2 菌株插入位置與方向一致，共偵測到 21 個插入點，其中 nt 2,163,352 插入點有 1 菌株覆蓋率低於 10% DP。

C00041 聚集中，有 1 株菌因 trimming 後序列讀長未達標準，無 IS6110 分析結果，其餘 10 株菌共偵測到 11 個插入點，插入方向皆相同，在 2 個相距 2 bp 的插入點分別有 7 株(nt 2,010,735)及 3 株(nt 2,010,727)帶有插入訊號；nt 2,165,962 有 2 株未偵測到插入訊號，其餘 8 株中有 7 株訊號覆蓋率低於 10% DP，

1 株為 11% DP。

C00042 聚集中，共偵測到 30 個插入點，插入方向皆相同，其中 mCDC031 為混合菌株，有 18 個插入點在其餘 3 菌株中無插入訊號，剩餘 12 個插入點中，nt 2,266,660 有 2 菌株無插入訊號，1 菌株覆蓋率小於 10% DP，nt 2,266,660 有 1 菌株無插入訊號。

C00048 聚集中，2 菌株插入位置與方向一致，共偵測到 13 個插入點。

C00049 聚集中，3 株菌皆帶有 11 個插入點的訊號，插入方向相同。

C00051 聚集中，3 株菌皆帶有 18 個插入點的訊號，插入方向相同。

C00065 聚集中，3 株菌共偵測 9 個插入點的訊號，插入方向皆一致，在 2 個相距 2 bp 的插入點分別有 2 株(nt 2,163,464)及 1 株(nt 2,163,466)帶有插入訊號。

C00070 聚集中，2 菌株插入位置與方向一致，共偵測到 21 個插入點。

C00074 聚集中，2 菌株插入位置與方向一致，共偵測到 8 個插入點。

C00113 聚集中，2 菌株插入位置與方向一致，共偵測到 19 個插入點，其中 nt 2,165,962 插入點 1 菌株覆蓋率低於 10% DP。

### 2.3 演化樹與傳播鏈分析

選取本次挑選菌株含有流行病學佐證資料、XDR 及多菌株之聚集進行演化樹與 core genome SNV 差異數分析。發現：

C00002 共有 3 名個案，分別來自彰化埤頭(2 名)、新北板橋(1 名)，彰化 2 分離菌株 core genome SNV 差異數為 4，彰化與新北菌株最小 SNV 差異數為 11 (圖二)。

C00009 共有 5 名個案，分別來自花蓮秀林(2 名)、花蓮市(1 名)、宜蘭大同

(1名)、台北士林(1名)，其中花蓮秀林的2名個案為同一社區居民，SNV 差異數為4，與花蓮市個案最小 SNV 差異數為4，3名花蓮個案與宜蘭大同個案最小 SNV 差異數為8、與台北市林個案最小 SNV 差異數為68(圖三)。

C00015 共有4名個案，皆來自雲林崙背鄉，其中3名為XDR，4名間 SNV 差異數為2至3(圖四)。

C00041 共有11名個案，分別來自花蓮市(2名)、花蓮萬榮(2名)、花蓮卓溪(1名)、花蓮秀林(1名)、花蓮吉安(1名)、屏東泰武(2名)、台南仁德(1名)及桃園大溪(1名)，其中屏東泰武2名個案與台南仁德1名個案間 SNV 差異數為0，屏東泰武2名個案為父子關係，此11名個案除桃園大溪(1名)及花蓮秀林(1名)SNV 最小差異數為12與23外，其餘9名個案 SNV 最小差異數在2至3之間(圖五)。

## (二) PacBio 定序

### 1. 菌株挑選：

自 MDR 聚集 C00041 選取 2012 年追管登記個案不同抗藥圖譜之分離菌 4 株，傳統藥敏皆為 RIF、INH、EMB 抗藥，1 株合併 SM、ETH 抗藥，1 株合併 SM、FQ 抗藥，1 株合併 FQ、ETH 抗藥，1 株合併 ETH 抗藥；Illumina 全基因定序分析分子抗藥結果，花蓮秀林菌株與 SM 傳統藥敏結果不一致，屏東泰武菌株則與 ETH 傳統藥敏結果不一致(表三十一)。

### 2. WGS 資料分析：

PacBio 定序 4 菌株，數據以 HGAP 演算進行基因組裝，再經環型基因體演算校正(circulator)，分別得 1、3、5、13 條基因重疊群(contigs)，基因體大小最大為花蓮萬榮菌株 mCDC108 4,389,983 bp (1 contig)，最小為花蓮吉安菌株



mCDC107 4,367,175 bp (13 contigs)，接近參考菌株 H37Rv 基因體 4,411,532 bp 大小，可達 98.99%以上(表三十二)。

以 1 contig 菌株 mCDC108 做為參考序列，比對該菌株 illumina 定序結果(圖六)，並分析 illumina 序列覆蓋深度低於 40%平均定序深度的區域，11 個低覆蓋深度區域有 8 個位於膜蛋白質 PE-PGRS 家族，1 個位在細胞毒性、宿主適應性相關的膜蛋白 mce1A、1 個位於功能尚未明朗的膜蛋白質 Rv2081c，1 個位在 PE-PGRS54 與功能應為 acetolactate synthase 的 ilvX 兩基因間區域(表三十三)。

#### 四、討論

##### 1. DNA 萃取方式評估：

WGS 定序的品質取決於所萃取基因體 DNA 的完整性，為避免 DNA 在萃取過程中斷裂與受損，應避免高溫、強酸強鹼、劇烈震盪、超音波或微珠破菌等製備步驟，受限於結核菌群需經去活化才可進行開放式分生操作，培養所收集的菌體選擇 80°C 加熱一小時去活化而非 95°C。本研究於 illumina 定序平台使用傳統 phenol/chloroform 法萃取 DNA 作為上機樣品，而在 PacBio 定序平台的部分，由於其 DNA 品質要求相較 Illumina 更為嚴苛，在本年度研究中選用商用試劑萃取 DNA 進行樣本製備(表三十四)。

##### 2. Illumina WGS 定序平台評估：

全基因體定序的平台與分析在生物學領域蓬勃發展，龐大的資訊量需要有系統地做整合以便於後續利用，結核菌全基因體的免費分析工具如 Mykrobe predictor、TB profiler、TGS-TB、PhyRes SE (表三十五)，提供了非生物資訊學背景的結核病研究人員介面簡單、單鍵上傳定序原始檔、快速獲得抗藥性與基因型分析結果的免安裝線上服務。各分析工具所提供的分析結果受限於所建立的藥物抗藥性基因資料庫完整性、基於不同方法學所使用的基因型分析演算方式，各有其侷限之處【21】。考量功能的完整性，本研究計畫選用 TGS-TB 分析抗藥性與基因型以及傳播鏈關係，其分析效能如下。

##### 2.1 抗藥性

本年度所挑選的 MDR 菌株以傳統藥物敏感性實驗為黃金準則，在 TB profiler 抗藥分析中，MDR 檢測敏感性可達 100%，優於現行疾管署一線藥物分子檢測之敏感性。

在檢測敏感性低的藥物中，SM 的檢測敏感性低於 80%。*gid* 基因被認為與 SM 低濃度抗藥相關，然而因有藥敏實驗佐證的位點資料不足，故在 SM 抗藥分析中，主要分析與抗藥高度相關的 *rrs* 與 *rpsL* 基因。本研究在 5 株 SM 抗藥但 *rrs*

與 *rpsL* 無相關抗藥位點的菌株中，發現 E170stop、V77G、nt 102 del G、nt 350 del G 等四種突變，nt 102 del G 在文獻中顯示抗藥關聯性不顯著【22】，nt 350 del G 尚未有藥敏資料，而其中 2 株帶有 V77G 的菌株在交差比對該聚集其它菌株後，顯示此位點可能與抗藥無關，而為該聚集之共同位點變異(圖七)。

而在 PZA 抗藥分析上，TB profiler 檢測敏感性則低於疾管署的檢驗效能，由於 *pncA* 其抗藥基因非熱區形式，驗證各種突變的抗藥性耗時且不易全面評估，因而現行疾管署的 PZA 分子抗藥以 *pncA* 是否發生啟動子區域突變，及蛋白質轉譯是否產生胺基酸突變做為分類準則敏感性較佳，而在後續研究中應建立與抗藥性無關聯性位點資料庫排除偽陽性菌株。

本研究選用 TB profiler 做為快速分析結核菌抗藥性的工具，其抗藥性分析整合目前最完整的位點資料庫，對混合菌株的抗藥位點偵測達 10% DP，可提供快速且高敏感性的分析結果【23】，然而其囊括的大量已發表文獻記載位點資料未經由完整的藥敏實驗驗證，使得某些菌株會產生偽陽性的檢測結果，如本次研究所發現的 *rrs a1401g* 與 SM 抗藥性的錯誤關聯，以及已證實為 EAI lineage marker 的 *kasA G312S* 被定義為 INH 抗藥相關。因此，在 TGS-TB 羅列的抗藥分析工具中，TB profiler 為提供參考之檢測結果，而 AMR prediction 所提供的位點為已有實際藥敏實驗評估抗藥關聯性之位點。

## 2.2 基因型

本研究所選用的菌株中，1 株為混合基因型菌株，在 spoligotyping 演算結果中有 3 個偵測位點實際實驗結果訊號偏弱，TGS-TB 所開發之全基因 spoligotyping 演算工具偵測混合基因型可達 10% DP，然而其顯示設計無法依定序訊號強度區分結果為單基因型或多基因型混合，而使此菌株的 ST no. 被判定為 undefined，然而 TB profiler 的 lineage 分析結果指出此菌株分別帶有 lineage1.2.1 Indo-Oceanic EAI2 與 lineage2.1 East-Asian 的辨識序列【24】，IS6110 插入點分析則顯示與同聚集的 EAI2-Manila 菌株相比，該菌株帶有 12 個相同的插入位點並多了 18 個不同

位點的插入訊號，綜合各工具的分析結果可判別此菌株為 EAI2 基因型混有 East-Asian 基因型。

### 2.3 演化樹與傳播鏈

演化樹與傳播鏈分析功能分析同聚集基因型菌株的演化關係與位點差異性，在英國全基因體分析結核病單核苷酸位點多型性(SNP)的研究中，近期感染的 SNP 差異數由 0 至 6 個不等【25-27】，本研究中 C00009 同社區個案 SNV 差異數為 4，C00041 的家庭群聚個案在 SNV 差異數為 0，與相關文獻結果吻合，於次年度的聚集菌株挑選規劃中，可望達成更充足的樣本分析結果，在未來定義群聚事件提供鑑別的分析準則。

另外，藉由交叉分析 C00041 聚集菌株演化樹與抗藥圖譜之關係，可發現抗藥位點隨演化樹分佈情形具有遞增性，在 2 株 SM 抗藥但尚未發現已知基因突變的菌株可見其演化關係與其它帶有 *rrs* a906g 較遠，其抗藥機制可能與現有已知之機轉不同(圖七)。

### 3. PacBio WGS 定序平台評估：

為評估 Illumina 定序效率不良區域是否影響抗藥性、基因型與演化傳播鏈分析，以 PacBio 基因組裝之 1 contig 做為參考序列，比對 illumina 同菌株之定序結果，定序覆蓋率低於 40% DP 區域皆分佈在膜蛋白質基因，其中 PE-PGRS 膜蛋白質家族佔半數以上，PE-PGRS 膜蛋白質與宿主免疫性相關，具有高基因重組率與高 GC%含量之特性【28-29】，大多數結核菌全基因體研究不將此高變異區列入序列分析，此結果亦顯示 PacBio 可克服 Illumina 定序效率不佳的高 GC%區域。

比較 4 株 C00041 聚集菌株的 PacBio 定序結果，總定序輸出量(total output)低於 300 M bp 的 2 菌株，基因組裝後最大 contig 無法達 90% H37Rv 基因體大小，而最大 contig 可達 99%以上 H37Rv 基因體大小的 2 菌株，總定序輸出量分別為

579.81 M bp (131 倍 H37Rv 基因體覆蓋深度)、479.8 M bp (108 倍 H37Rv 基因體覆蓋深度)(表三十二)，顯示以現階段 DNA 萃取至上機流程，大於 100 倍參考菌株基因體大小的定序輸出量基因組裝效率較佳。

本年度 PacBio 單次定序上機量約為 14  $\mu$ g DNA，考量結核菌細胞壁結構組成不易破菌，且樣本應極力避免 DNA 斷裂以達長讀長定序之最佳優勢，後續將測試不同 DNA 萃取方式，如 DNA 萃取長度可達 30 Kb 的 Isoplant 試劑組(日本和光)，以優化 PacBio 定序效能。

## 五、結論與建議

因應結核菌高 GC 含量與高重複片段的基因體組成及特性，本研究在 illumina 定序平台上，使用物理性的 DNA 片段化流程製備樣本庫，避免酵素型片段化流程使定序效能高 GC 含量序列不佳，並選用高定序輸出量的 HiSeq 平台設定 240 倍的定序深度，以求完整取得結核菌的全基因體序列。根據已完成 TGS-TB 分析的 56 筆 MDR 菌株結果可知，使用 phenol/chloroform 萃取的 DNA 樣品，paired-end 定序深度大於 100 倍時，比對參考序列 H37Rv 的覆蓋率皆大於 98%，然而部分菌株因 trimming 後長度不足，使基因型分析受限，故**建議未來之上機模式可將定序深度降至 120 倍，定序讀長提升至 300 bases。**

現有策略可有效鑑別 10 種藥物(INH、RIF、PZA、SM、EMB、ETH、FQ、AMK、CAP、KM)已知的抗藥基因相關突變位點，然而在突變位點抗藥關聯性上，應建立帶有傳統藥敏資料佐證的位點資料庫，以排除可能造成偽陰性或偽陽性判斷的突變位點。另外，分子與傳統藥敏不一致菌株可進一步做為未來尋找未知抗藥相關基因體變異的候選菌株。**建議未來可合併演化分析結果，挑選抗藥機制未明的菌株進行更深入的藥物敏感性與基因體分析。**

TGS-TB 在分析結核菌基因型上，單一基因型菌株可提供與 spoligotyping 實驗一致的基因型分析結果，而在混合不同基因型的菌株中，由於 spoligotype 為混合的訊號無法判別正確的 ST no.，故**建議 WGS 分析，宜應以單一菌落或單一菌群之培養做為分析樣品。**

而在演化關係分析中，core genome SNV 分析在 2 名同社區個案得到 SNV 差異數 4，2 名家庭個案 SNV 差異數 0 的結果，與研究文獻提供的群聚事件 SNV 值落在 0 至 6 之間吻合，當缺乏流行病學資料佐證調查時可提供更詳細的傳染鏈資訊以正確及精準追蹤感染途徑。而在 PacBio 平行對照的定序結果顯示，illumina 的定序效能低的區域落在與結核菌引發宿主免疫反應相關的膜蛋白質家族 PE-PGRS，該蛋白質家族因高變異性與高 GC 含量使得定序分析不易，因而多自全基因序列分析中隔離，在評估 illumina 應用於結核菌抗藥性、基因行與演化分析上，提供了正面的資訊。**建議積極規劃全基因定序抗藥性**

**與基因型之資料架構，以串連檢驗結果至傳染病通報與動態即時監測系統。**

全基因定序的發展成熟與成本下降使得此技術在臨床檢驗的應用上日益廣泛，Illumina 在病原菌檢測的應用上，可提供大量抗藥性與基因型資訊，提供臨床端一個完整的基因體分析結果，PacBio 則可克服 Illumina 高 GC 含量序列定序性能不良，以及短讀長定序技術限制而遺漏的大片段基因變異【30】，在建立不同基因型菌株的參考序列資料庫上，提供了良好的技術平台，然而受限於菌株增量培養耗時、菌株 DNA 檢體處理流程繁瑣、儀器昂貴、操作人員須經專業訓練、全基因定序價格相較一般臨床檢驗仍所費不貲，在後續大量應用於公共衛生與臨床檢驗上，朝樣本處理簡單化與分析自動化為本技術的發展趨勢。英國 2015 年發表將 Oxford Nanopore 掌上型定序儀 MinION 定序技術直接應用在檢驗痰檢體的結核菌的全基因體【31】，在 2017 年評估該技術應用於抹片陽性痰檢體當日診斷以及傳染疫情監測的檢測，成本最低可達台幣約 4000 元【32】，該技術定序讀長平均 1 kb，介於 Illumina 與 PacBio 之間，最低所需 DNA 上機量 200 ng 與 Illumina 相當，具有樣本庫製備流程簡單化、分析自動化與可攜式等優勢，雖然最低菌量偵測極限尚不及 Xpert MTB/RIF (LOD: 133 CFU/ml，約 0.6 pg DNA/ml)，仍為具有整合進臨床檢驗流程潛力之技術平台，惟因在台灣尚未引進，**建議未來上市可將 nanopore 做為另一評估之標的。**

## 六、計畫重要研究成果及具體建議

### (一) 成果

#### 第一年

1. 完成 Illumina 全基因定序結核菌之 DNA 萃取、樣本庫置備、資料分析的流程建立。
2. 確認 10 種 TB 治療藥物抗藥性分析之準確性，並提供尚待確認之可能相關機轉之訊息。
3. 以台灣盛行之 MDR 與 XDR 聚集菌株評估線上結核菌全基因定序分析工具之效能，在抗藥性、基因型以及群聚傳染源分析上，可提供相比現行疾管署檢驗項目更完整與詳細的資訊。
4. 完成 PacBio 全基因定序測試，取得台灣地區高盛行 MDR 聚集基因型之參考序列，可做為優化定序流程，與建立各基因型參考序列之基礎。

### (二) 具體建議

1. 配合分析平台之設定，更改 Illumina 讀長及定序深度，以符合經濟效益與分析需求。
2. 規劃全基因定序抗藥性與基因型之資料架構，以串連檢驗結果至傳染病通報與動態即時監測系統。
3. 評估其它極具臨床檢驗應用性的簡式全基因定序平台，以縮短檢測時效。



## 七、參考文獻

依一般科學論文之參考文獻撰寫方式，列出所引用之參考文獻，並於計畫內容引用處標註之。

1. Global tuberculosis report 2015. <http://who.int/tb/publications/en/>
2. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CL, Bowden R, Drobniowski FA, Allix-Béguec C, Gaudin C, Parkhill J, Diel R, Supply P, Crook DW, Smith EG, Walker AS, Ismail N, Niemann S, Peto TE; Modernizing Medical Microbiology (MMM) Informatics Group. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis*. 2015 Oct;15(10):1193-202.
3. Köser CU, Bryant JM, Becq J, Török ME, Ellington MJ, Marti-Renom MA, Carmichael AJ, Parkhill J, Smith GP, Peacock SJ. Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. *N Engl J Med*. 2013 Jul 18;369(3):290-2.
4. Wang F, Shao L, Fan X, Shen Y, Diao N, Jin J, Sun F, Wu J, Chen J, Weng X, Cheng X, Zhang Y, Zhang W. Evolution and transmission patterns of extensively drug-resistant tuberculosis in China. *Antimicrob Agents Chemother*. 2015 Feb;59(2):818-25.
5. Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida DV, Alvarado L, Chapman SB, Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S, Howarth C, Larimer JD, Maharaj K, Pearson MD, Priest ME, Zeng Q, Padayatchi N, Grosset J, Young SK, Wortman J, Mlisana KP, O'Donnell MR, Birren BW, Bishai WR, Pym AS, Earl AM. Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of *Mycobacterium tuberculosis* Isolates from KwaZulu-Natal. *PLoS Med*. 2015 Sep 29;12(9):e1001880.
6. Schurch, A.C., Kremer, K., Daviena, O., Kiers, A., Boeree, M.J., Siezen, R.J., van Soolingen, D. High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *J Clin Microbiol*. 2010;48:3403–3406
7. Walker, T.M., Ip, C.L., Harrell, R.H., Evans, J.T., Kapatai, G., Dedicoat, M.J., Eyre, D.W., Wilson, D.J., Hawkey, P.M., Crook, D.W., Parkhill, J., Harris, D., Walker, A.S., Bowden, R., Monk, P., Smith, E.G., Peto, T.E. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis*. 2013;13:137–146

8. Gardy, J.L., Johnston, J.C., Ho Sui, S.J., Cook, V.J., Shah, L., Brodtkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., Varhol, R., Birol, I., Lem, M., Sharma, M.K., Elwood, K., Jones, S.J., Brinkman, F.S., Brunham, R.C., Tang, P. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med.* 2011;364:730–739
9. Bryant JM, Schürch AC, van Deutekom H, Harris SR, de Beer JL, de Jager V, Kremer K, van Hijum SA, Siezen RJ, Borgdorff M, Bentley SD, Parkhill J, van Soolingen D. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect Dis.* 2013 Feb 27;13:110.
10. Roetzer, A., Diel, R., Kohl, T.A., Ruckert, C., Nubel, U., Blom, J., Wirth, T., Jaenicke, S., Schuback, S., Rusch-Gerdes, S., Supply, P., Kalinowski, J., Niemann, S. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med.* 2013;10:e1001387
11. Kohl TA, Diel R, Harmsen D, Rothgänger J, Walter KM, Merker M, Weniger T, Niemann S. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J Clin Microbiol.* 2014 Jul;52(7):2479-86.
12. Kato-Maeda M, Ho C, Passarelli B, Banaei N, Grinsdale J, Flores L, Anderson J, Murray M, Rose G, Kawamura LM, Pourmand N, Tariq MA, Gagneux S, Hopewell PC. Use of whole genome sequencing to determine the microevolution of *Mycobacterium tuberculosis* during an outbreak. *PLoS One.* 2013;8(3):e58235.
13. Guidelines for surveillance of drug resistance in tuberculosis - 5th edition [http://who.int/tb/publications/2015/drs\\_guidelines/en/](http://who.int/tb/publications/2015/drs_guidelines/en/)
14. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet.* 2011 Dec 18;44(1):106-10.
15. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW; COMPASS-TB Study Group. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *Lancet Respir Med.* 2016 Jan;4(1):49-58.
16. Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733 (2015).

17. Sekizuka, T. et al. TGS-TB: Total Genotyping Solution for Mycobacterium tuberculosis Using Short-Read Whole-Genome Sequencing. *PloS One* 10, e0142951 (2015).
18. Spies, F. S. et al. Streptomycin Resistance and Lineage-Specific Polymorphisms in Mycobacterium tuberculosis gidB Gene ▽. *J. Clin. Microbiol.* 49, 2625–2630 (2011).
19. Sun, Y.-J., Lee, A. S. G., Wong, S.-Y. & Paton, N. I. Analysis of the role of Mycobacterium tuberculosis kasA gene mutations in isoniazid resistance. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* 13, 833–835 (2007).
20. Domínguez, J. et al. Clinical implications of molecular drug resistance testing for Mycobacterium tuberculosis: a TBNET/RESIST-TB consensus statement. *Int. J. Tuberc. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis.* 20, 24–42 (2016).
21. Satta, G., Atzeni, A. & McHugh, T. D. Mycobacterium tuberculosis and whole genome sequencing: a practical guide and online tools available for the clinical microbiologist. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* 23, 69–72 (2017).
22. Wong, S. Y. et al. Mutations in gidB Confer Low-Level Streptomycin Resistance in Mycobacterium tuberculosis. *Antimicrob. Agents Chemother.* 55, 2515–2522 (2011).
23. Coll, F. et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med.* (2015). doi:10.1186/s13073-015-0164-0
24. Coll, F. et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. *Nat. Commun.* 5, 4812 (2014).
25. Walker, T. M. et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. *Lancet Infect. Dis.* 13, 137–146 (2013).
26. Casali, N. et al. Whole Genome Sequence Analysis of a Large Isoniazid-Resistant Tuberculosis Outbreak in London: A Retrospective Observational Study. *PLoS Med.* 13, (2016).
27. Nikolayevskyy, V., Kranzer, K., Niemann, S. & Drobniowski, F. Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks: A systematic review. *Tuberc. Edinb. Scotl.* 98, 77–85 (2016).
28. Banu, S. et al. Are the PE-PGRS proteins of Mycobacterium tuberculosis variable surface antigens? *Mol. Microbiol.* 44, 9–19 (2002).
29. Espitia, C. et al. The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a new family of fibronectin-binding proteins? *Microbiol. Read. Engl.* 145 ( Pt 12), 3487–3495 (1999).

30. Shitikov, E. A. et al. Unusual large-scale chromosomal rearrangements in *Mycobacterium tuberculosis* Beijing B0/W148 cluster isolates. *PloS One* 9, e84971 (2014).
31. Brown, A. C. et al. Rapid Whole-Genome Sequencing of *Mycobacterium tuberculosis* Isolates Directly from Clinical Samples. *J. Clin. Microbiol.* 53, 2230–2237 (2015).
32. Votintseva, A. A. et al. Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. *J. Clin. Microbiol.* 55, 1285–1298 (2017).

## 八、圖、表

表一、本計畫評估之第二代與第三代全基因定序平台

Platform	Read length	Advantage	Disadvantage	DNA submission
Illumina HiSeq	250 bp x 2	<ol style="list-style-type: none"> <li>1. <b>High throughput</b></li> <li>2. For resequencing (reference genome mapping)</li> </ol>	<ol style="list-style-type: none"> <li>1. Short read length: <b>hard to overcome the repeated sequence</b> (ex: IS6110, 1.3 kb) or DNA rearrangement</li> </ol>	>200 ng
PacBio SMRT	>10 Kb	<ol style="list-style-type: none"> <li>1. <b>Long read length</b></li> <li>2. For <i>de Novo</i> assembly sequencing</li> <li>3. Lack of GC% or sequence complexity bias</li> <li>4. Epigenome characterization</li> </ol>	<ol style="list-style-type: none"> <li>1. High quality DNA samples are required (purity, amount, and integrity)</li> <li>2. Higher cost</li> </ol>	>14 µg

表二、挑選 2009-2014 抗藥性聚集菌株

TCDC Cluster No.	Lineage	ST No.	MDR 菌株	XDR 菌株
C00041	Beijing	1	11*	
C00017	Beijing	1	5	1
C00015	Beijing	1	1	3
C00113	Beijing	1		2
C00009	H3	316	5	
C00042	EAI2-Manila	19	3	
	EAI2-Manila(mix)	19	1	
C00065	Beijing	250	3	
C00049	H3	50	3	
C00051	H	742	3	
C00002	H3	2090	3	
C00024	Beijing	190	2	
C00037	Manu ancestor	523	2	
C00003	H3	2090	2	
C00048	T1	53	2	
C00074	T2	52	2	
C00070	Unknown	1487	2	

\*挑選 4 株 C00041 MDR 菌株進行 PacBio 定序。

表三、TGS-TB 全基因定序抗藥性分析與傳統藥敏結果比較

Drugs	TB profiler	Conventional DST		Total	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
		Resistant	Susceptible					
Rifampin (RIF)	Resistant	56	0	56	100.0	-	100.0	-
	Susceptible	0	0					
Isoniazid (INH)	Resistant	56	0	56	100.0	-	100.0	-
	Susceptible	0	0					
Ethambutol (EMB)	Resistant	25	5	56	100.0	83.9	83.3	100.0
	Susceptible	0	26					
Streptomycin (SM)	Resistant	19	1	56	79.2	96.9	95.0	86.1
	Susceptible	5	31					
Capreomycin (CAP)	Resistant	5	2	56	83.3	96.0	71.4	98.0
	Susceptible	1	48					
Kanamycin (KM)	Resistant	7	3	56	100.0	93.9	70.0	100.0
	Susceptible	0	46					
Amikacin (AM)	Resistant	7	0	56	100.0	100.0	100.0	100.0
	Susceptible	0	49					
Fluoroquinolone (FQ)	Resistant	21	0	56	100.0	100.0	100.0	100.0
	Susceptible	0	35					
Ethionamide (ETH)	Resistant	23	13	56	85.2	55.2	63.9	80.0
	Susceptible	4	16					
Pyrazinamide (PZA)	Resistant	13	0	56	81.3	100.0	100.0	93.0
	Susceptible	3	40					

表四、TB profiler 提供抗藥基因分析與疾管署分子抗藥檢測服務比較

Drugs	TB profiler DR genes	疾管署提供之一、二線抗藥基因定序
RIF	<i>rpoB</i>	codon 332-494 (hotspot region: 426-452)
	<i>rpoC</i>	
INH	<i>inhA</i>	(即 inhAr) <i>fabG1</i> nt -230~+145 codon 228-472
	<i>fabG1</i> promoter	
	<i>katG</i>	
	<i>kasA</i>	
ETH	<i>ahpC</i>	<i>inhA</i> <i>fabG1</i> promoter <i>ethA</i> <i>ethR</i>
	<i>inhA</i>	
	<i>ethA</i>	
	<i>ethR</i>	
EMB	<i>embB</i>	<i>embB</i> <i>embA</i> <i>embC</i> <i>embR</i>
	<i>embA</i>	
	<i>embC</i>	
	<i>embR</i>	
SM	<i>rrs</i>	nt 1330-1520
	<i>rpsL</i>	
CAP	<i>rrs</i>	nt 1330-1520
	<i>tlyA</i>	
AM	<i>rrs</i>	nt 1330-1520
KM	<i>rrs</i>	nt 1330-1520
	<i>eis</i> promoter	nt -325 ~ +40
OFX	<i>gyrA</i>	codon 18-128 (hotspot: codon 88-94)
	<i>gyrB</i>	codon 390-558 (hotspot: codon 461-499)
PZA	<i>pncA</i>	promoter nt -67~-1 +FL coding sequence

表五、疾管署分子抗藥檢測分析與傳統藥敏結果比較

Drugs	CDC DR seq	Conventional DST		Total	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
		Resistant	Susceptible					
Rifampin (RIF)	Resistant Susceptible	55 1	0 0	56	98.2	-	100.0	-
Isoniazid (INH)	Resistant Susceptible	54 2	0 0	56	96.4	-	100.0	-
Capreomycin (CAP)	Resistant Susceptible	5 1	2 48	56	83.3	96.0	71.4	98.0
Kanamycin (KM)	Resistant Susceptible	7 0	3 46	56	100.0	93.9	70.0	100.0
Amikacin (AM)	Resistant Susceptible	7 0	0 49	56	100.0	100.0	100.0	100.0
Fluoroquinolone (FQ)	Resistant Susceptible	21 0	0 35	56	100.0	100.0	100.0	100.0
Ethionamide (ETH)	Resistant Susceptible	23 4	13 16	56	85.2	55.2	63.9	80.0
Pyrazinamide (PZA)	Resistant Susceptible	16 0	0 40	56	100.0	100.0	100.0	100.0



表六、RIF 抗藥位點統計

RIF	R (%)	<i>E. coli</i> codon
<i>rpoB</i> D435Y	3 (5.4)	D516Y
<i>rpoB</i> D435V	2 (3.6)	D516V
<i>rpoB</i> Q432K	2 (3.6)	Q513K
<i>rpoB</i> H445R	2 (3.6)	H526R
<i>rpoB</i> L452P	5 (8.9)	L533P
<i>rpoB</i> S450L	41 (73.2)	S531L
<i>rpoB</i> V170F	1 (1.8)	V176F

表七、INH 抗藥位點統計

INH	R (%)
<i>fabG1</i> c-15t	16 (28.6)
<i>fabG1</i> c-15t, <i>inhA</i> I194T	3 (5.4)
<i>fabG1</i> c-15t, <i>inhA</i> I194T, <i>kasA</i> G312S	4 (7.1)
<i>fabG1</i> c-15t, <i>katG</i> A106V	1 (1.8)
<i>fabG1</i> c-15t, <i>katG</i> S315T	10 (17.9)
<i>fabG1</i> t-8c	2 (3.6)
<i>katG</i> D735A	2 (3.6)
<i>katG</i> S315T	18 (32.1)

表八、EMB 抗藥位點統計

EMB	R (%)	S (%)
<i>embA</i> c-11a	2 (3.6)	
<i>embA</i> c-12t, <i>embB</i> G406S	2 (3.6)	1 (1.8)
<i>embB</i> D1024N	1 (1.8)	
<i>embB</i> Q497K	2 (3.6)	1 (1.8)
<i>embB</i> Q497P D1024N	9 (16.1)	
<i>embB</i> G406D		1 (1.8)
<i>embB</i> M306L	1 (1.8)	1 (1.8)
<i>embB</i> M306V	8 (14.3)	1 (1.8)
<i>embB</i> Y319S	1 (1.8)	
no TB profiler defined DR mutations		25 (44.6)

表九、SM 抗藥位點統計

SM	R (%)	S (%)
<i>rpsL</i> K43R	7 (12.5)	
<i>rpsL</i> K43R, <i>rrs</i> a1401g	6 (10.7)	
<i>rrs</i> a1401g		1 (1.8)
<i>rrs</i> a906g	6 (10.7)	
no TB profiler DR mutations	<b>5 (8.9)</b>	31 (55.4)
<i>gid</i> E170Stop, nt 555_556 ins T	1 (1.8)	
<i>gid</i> V77G E92D	2 (3.6)	
<i>gid</i> nt 102 del G (50%), E92D	1 (1.8)	
<i>gid</i> nt 350 del G	1 (1.8)	

表十、二線針劑藥物 CAP、KM、AM 抗藥位點統計

CAP	R (%)	S (%)
<i>rrs</i> a1401g	5 (8.9)	2 (3.6)
no TB profiler defined DR mutations	1 (1.8)	48 (85.7)
KM	R (%)	S (%)
<i>eis</i> g-10c		3 (5.4)
<i>rrs</i> a1401g	7 (12.5)	
no TB profiler DR mutations		46 (82.1)
AM	R (%)	S (%)
<i>rrs</i> a1401g	7 (12.5)	
no TB profiler defined DR mutations		49 (87.5)

表十一、FQ 抗藥位點統計

FQ	R (%)	S (%)
<i>gyrA</i> D94A	2 (3.6)	
<i>gyrA</i> D94G	15 (26.8)	
<i>gyrA</i> D94H	1 (1.8)	
<i>gyrA</i> D94Y	3 (5.4)	
no TB profiler defined DR mutations		35 (62.5)

表十二、ETH 抗藥位點統計

ETH	R (%)	S (%)
<i>fabG1</i> c-15t	18 (32.1)	9 (16.1)
<i>fabG1</i> c-15t, <i>inhA</i> I194T	5 (8.9)	2 (3.6)
<i>fabG1</i> t-8c		2 (3.6)
no TB profiler defined DR mutations	<b>4 (7.1)</b>	16 (28.6)
<i>ethA</i> nt 671-672 ins GC	1 (1.8)	
<i>ethA</i> F66C	1 (1.8)	
<i>ethA</i> Q254P	1 (1.8)	
<i>ethA</i> wt, <i>ethR</i> wt	1 (1.8)	

表十三、PZA 抗藥位點統計

PZA	R (%)	S (%)
<i>pncA</i> C14R	4 (7.1)	
<i>pncA</i> I133T	2 (3.6)	
<i>pncA</i> L19R	1 (1.8)	
<i>pncA</i> T-11C	2 (3.6)	
<i>pncA</i> T135P	1 (1.8)	
<i>pncA</i> T76P	3 (5.4)	
no TB profiler defined DR mutations	<b>3 (5.4)</b>	40 (71.4)
<i>pncA</i> nt 292_293 ins A	2 (3.6)	
<i>pncA</i> L120R	1 (1.8)	





表十五、C00002 聚集菌株 TGS-TB IS6110 插入位置分析表

isolates Position on H37Rv	mCDC003		mCDC004		mCDC005	
	Count	F/R	Count	F/R	Count	F/R
1893	53	R	39	R	49	R
850168	60	R	44	R	58	R
1074735	39	R	44	R	36	R
1175644	58	F	46	F	37	F
1998657	32	F	39	F	35	F
1998792	58	F	38	F	40	F
2163464	42	F	41	F	58	F
3120049	62	R	39	R	44	R
3310456	53	R	39	R	43	R
3554820	61	R	48	R	31	R
3750540	33	F	44	F	59	F

F/R: forward or reverse, insertion direction

表十六、C00003 聚集菌株 TGS-TB IS6110 插入位置分析表

isolates DP Position on H37Rv	mCDC047 108.3×		mCDC048 108.5×	
	Count	F/R	Count	F/R
888786	30	F	62	F
888832	49	F	54	F
1074735	53	R	31	R
1982241	37	R	25	R
1998657	18	F	22	F
2040775	63	F	60	F
2163464	50	F	57	F
3120049	38	R	39	R
3310456	34	R	54	R
3554820	63	R	38	R
3750540	53	F	28	F

F/R: forward or reverse, insertion direction

表十七、C00009 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC025		mCDC026		mCDC027		mCDC028		mCDC029	
	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R
888786	49	F	43	F	43	F	45	F	50	F
888871	37	R	45	R	47	R	36	R	36	R
1024013	36	F	38	F	41	F	33	F	31	F
1074735	23	R	31	R	25	R	26	R	33	R
1998657	24	F	23	F	18	F	32	F	18	F
2010679	44	F	57	F	70	F	57	F	52	F
2163464	34	F	37	F	47	F	48	F	46	F
3120049	40	R	30	R	42	R	52	R	38	R
3123045	56	F	48	F	58	F	41	F	ND	
3123179	64	F	51	F	65	F	46	F	36	F
3310456	47	R	67	R	45	R	39	R	43	R
3750540	42	F	35	F	46	F	45	F	45	F

F/R: forward or reverse, insertion direction, ND: not detected

表十八、C00015 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC006		mCDC007		mCDC008		mCDC009	
	Count	F/R	Count	F/R	Count	F/R	Count	F/R
1591	57	F	53	F	39	F	41	F
888989	47	F	56	F	44	F	46	F
1168747	51	F	63	F	50	F	39	F
1262964	43	R	44	R	46	R	33	R
1543973	46	R	56	R	30	R	28	R
1657016	56	R	45	R	51	R	38	R
1886327	33	R	57	R	45	R	50	R
1986639	45	R	61	R	44	R	49	R
2165962	14	R	<b>8</b>	R	<b>6</b>	R	<b>6</b>	R
2263628	69	R	55	R	43	R	43	R
2366891	54	F	46	F	50	F	73	F
2634050	48	F	41	F	49	F	44	F
3127931	46	F	54	F	53	F	56	F
3378554	58	R	78	R	42	R	52	R
3379024	55	F	56	F	52	F	53	F
3493911	48	R	45	R	49	R	55	R
3549196	22	F	35	F	45	F	45	F
3797824	58	F	51	F	51	F	65	F
3844682	49	R	36	R	53	R	32	R
4174241	49	F	49	F	58	F	39	F

F/R: forward or reverse, insertion direction, **Bold**: count/avg. depth<10%

表十九、C00017 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC001		mCDC020		mCDC021		mCDC022		mCDC023		mCDC024	
	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R
1591	44	F	39	F	42	F	40	F	43	F	34	F
888989	60	F	51	F	33	F	44	F	51	F	46	F
1262964	30	R	38	R	45	R	43	R	43	R	33	R
1543973	41	R	41	R	43	R	31	R	58	R	49	R
1657016	49	R	49	R	35	R	29	R	37	R	53	R
1986639	57	R	40	R	63	R	46	R	42	R	39	R
2040780	44	R	44	R	58	R	42	R	54	R	59	R
2165962	<b>8</b>	R	<b>9</b>	R	15	R	12	R	<b>10</b>	R	15	R
2263628	54	R	48	R	37	R	48	R	53	R	41	R
2274412	52	R	49	R	43	R	56	R	47	R	49	R
2636954	27	F	26	F	28	F	31	F	34	F	23	F
2639539	<b>8</b>	F	ND		<b>6</b>	F	ND		<b>8</b>	F	<b>7</b>	F
3127931	38	F	50	F	38	F	46	F	54	F	43	F
3378554	52	R	58	R	40	R	51	R	57	R	56	R
3494460	43	F	32	F	40	F	54	F	45	F	46	F
3549197	49	R	ND		ND		ND		ND		ND	
3549200	ND		39	R	32	R	53	R	61	R	46	R
3797824	60	F	35	F	51	F	38	F	62	F	45	F
3844682	42	R	39	R	45	R	51	R	46	R	49	R
4319652	ND		ND		ND		ND		<b>8</b>	F	ND	

F/R: forward or reverse, insertion direction, **Bold**: count/avg. depth<10%, ND: not detected

表二十、C00024 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC053		mCDC054	
	Count	F/R	Count	F/R
1591	38	F	40	F
888989	64	F	58	F
1262964	48	R	54	R
1543972	33	R	42	R
1657016	57	R	52	R
1986639	70	R	59	R
2164716	52	F	49	F
2165962	<b>10</b>	R	ND	
2263628	46	R	39	R
2366891	43	F	52	F
2367678	47	R	53	R
2634021	36	F	64	F
3127931	45	F	57	F
3379024	30	F	50	F
3493911	70	R	44	R
3547341	45	F	42	F
3549200	43	R	38	R
3797824	41	F	47	F
3844682	43	R	43	R

F/R: forward or reverse, insertion direction,  
**Bold**: count/DP<10%, ND: not detected



表二十一、C00037 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC036		mCDC037	
	Count	F/R	Count	F/R
1591	49	F	47	F
888786	42	F	54	F
1168965	36	F	47	F
1987570	59	R	53	R
2129600	49	F	58	F
2163352	12	F	<b>10</b>	F
2163389	20	F	26	F
2163458	21	F	16	F
2174808	52	R	51	R
2202073	55	F	54	F
2270536	40	F	39	F
2367198	34	R	64	R
2610979	43	R	41	R
2634349	39	F	45	F
3120524	43	R	52	R
3125127	40	R	45	R
3378554	52	R	56	R
3545553	43	F	55	F
3555643	56	F	58	F
3841801	46	R	54	R
4174337	53	R	56	R

F/R: forward or reverse, insertion direction,

**Bold:** count/avg. depth<10%

表二十二、C00041 聚集菌株 TGS-TB IS6110 插入位置分析

isolates	mCDC012		mCDC013		mCDC014		mCDC015		mCDC016		mCDC017		mCDC018		mCDC019		mCDC107		mCDC108		mCDC104
Position on H37Rv	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R	Count	F/R	input too short
1591	57	F	58	F	43	F	39	F	57	F	49	F	50	F	40	F	59	F	76	F	
850610	49	F	38	F	45	F	49	F	45	F	59	F	52	F	44	F	45	F	63	F	
1986639	51	R	41	R	48	R	57	R	50	R	38	R	57	R	56	R	60	R	49	R	
2010735	ND		57	F	63	F	ND		56	F	46	F	ND		44	F	49	F	51	F	
2010737	80	F	ND		ND		43	F	ND		ND		60	F	ND		ND		ND		
2165962	<b>6</b>	R	ND		<b>8</b>	R	12	R	<b>10</b>	R	<b>10</b>	R	<b>6</b>	R	<b>9</b>	R	ND		<b>10</b>	R	
2263779	45	R	50	R	43	R	58	R	46	R	41	R	42	R	62	R	63	R	66	R	
3127931	69	F	60	F	65	F	55	F	44	F	51	F	35	F	39	F	57	F	58	F	
3304688	55	R	64	R	43	R	58	R	48	R	43	R	53	R	58	R	47	R	56	R	
3378554	52	R	50	R	61	R	55	R	67	R	51	R	53	R	42	R	55	R	62	R	
3796688	32	F	47	F	41	F	57	F	41	F	54	F	65	F	48	F	55	F	63	F	

F/R: forward or reverse, insertion direction, **Bold**: count/avg. depth<10%, ND: not detected

表二十三、C00042 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC030		mCDC032		mCDC033		mCDC031	
	Count	F/R	Count	F/R	Count	F/R	Count	F/R
1591							12	F
888789							25	F
888870	60	R	46	R	46	R	24	R
932155	65	F	57	F	69	F	43	F
1168965							23	F
1721310	50	F	47	F	54	F	44	F
1946981	61	R	39	R	30	R	29	R
1987570							9	R
2038820	51	R	46	R	56	R	31	R
2129600							15	F
2163393							7	R
2163462							6	R
2174808							12	R
2266660	13	R	ND		6	R	ND	
2405499	47	F	53	F	40	F	21	F
2610979							16	R
2634351							14	F
3114330							13	R
3119184							19	R
3120524	63	R	46	R	34	R	25	R
3125125							22	R
3302966							25	F
3378554							19	R
3487912							26	F
3548562	46	R	ND		54	R	21	R
3551070	68	R	44	R	63	R	31	R
3554173	66	R	53	R	56	R	14	R
3555643							12	F
3841801							25	R
4230306	87	R	43	R	54	R	27	R

F/R: forward or reverse, insertion direction, **Bold**: count/avg. depth<10%,  
 ND: not detected, Box: only in mCDC031(mix-clone)

表二十四、C00048 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC051		mCDC052	
	Count	F/R	Count	F/R
1879	46	R	73	R
1074735	44	R	38	R
1243231	47	R	43	R
1851882	59	F	43	F
1998657	43	F	34	F
2163464	40	F	48	F
2165980	57	R	45	R
2263666	48	R	51	R
2579596	49	F	55	F
3120524	42	R	54	R
3310456	44	R	43	R
3750540	51	F	48	F
3796759	68	R	54	R

F/R: forward or reverse, insertion direction

表二十五、C00049 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC041		mCDC042		mCDC043	
	Count	F/R	Count	F/R	Count	F/R
1074735	37	R	27	R	20	R
1998657	31	F	25	F	35	F
2047364	34	R	51	R	45	R
2163464	55	F	45	F	27	F
2199705	34	R	39	R	40	R
2345884	40	R	47	R	40	R
3120049	31	R	43	R	35	R
3120524	42	R	39	R	40	R
3310456	43	R	37	R	22	R
3550867	49	F	27	F	48	F
3750540	36	F	35	F	45	F

F/R: forward or reverse, insertion direction

表二十六、C00051 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC044		mCDC045		mCDC046	
	Count	F/R	Count	F/R	Count	F/R
1591	46	F	38	F	52	F
475853	56	R	45	R	52	R
850602	55	R	60	R	49	R
1543973	42	R	42	R	38	R
1659185	39	R	42	R	40	R
1744347	48	F	63	F	26	F
1986639	57	R	59	R	42	R
2046901	48	R	37	R	43	R
2166244	52	F	55	F	52	F
2263635	45	F	46	F	42	F
2633605	51	F	44	F	24	F
2732424	57	R	64	R	44	R
2732673	53	R	48	R	57	R
3127931	55	F	52	F	49	F
3378554	59	R	50	R	53	R
3552676	51	F	47	F	68	F
3594427	51	R	58	R	50	R
3797824	56	F	70	F	50	F

F/R: forward or reverse, insertion direction

表二十七、C00065 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC038		mCDC039		mCDC040	
	Count	F/R	Count	F/R	Count	F/R
1074735	19	R	32	R	35	R
1543462	67	R	39	R	44	R
1998657	22	F	33	F	31	F
2041738	47	F	33	F	43	F
2163464	43	F	62	F	ND	
2163466	ND		ND		59	F
3120049	50	R	45	R	43	R
3310456	55	R	46	R	40	R
3750540	53	F	39	F	38	F

F/R: forward or reverse, insertion direction, ND: not detected

表二十八、C00070 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC034		mCDC035	
	Count	F/R	Count	F/R
1591	51	F	40	F
759412	47	R	44	R
888786	56	F	44	F
888790	55	R	55	R
1168965	47	F	30	F
1987120	39	R	57	R
2038819	42	R	61	R
2129600	46	F	51	F
2163352	11	F	<b>9</b>	F
2163421	14	F	20	F
2163470	11	F	16	F
2174808	49	R	51	R
2368531	47	R	60	R
2610979	34	R	46	R
2634349	39	F	45	F
3120524	55	R	38	R
3378554	65	R	46	R
3491533	52	R	38	R
3555643	47	F	59	F
3790369	51	R	54	R
3841801	57	R	54	R

F/R: forward or reverse, insertion direction,  
**Bold**: count/avg. depth<10%

表二十九、C00074 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC049		mCDC050	
	Count	F/R	Count	F/R
1527046	52	R	38	R
1976332	55	F	51	F
1981699	46	F	44	F
2010726	55	F	42	F
2041753	30	R	45	R
2751571	35	R	52	R
3120524	50	R	47	R
3709677	54	R	51	R

F/R: forward or reverse, insertion direction

表三十、C00113 聚集菌株 TGS-TB IS6110 插入位置分析

isolates Position on H37Rv	mCDC010		mCDC011	
	Count	F/R	Count	F/R
1591	47	F	44	F
1892	55	R	44	R
888989	63	F	39	F
1262964	45	R	30	R
1543973	44	R	41	R
1657016	58	R	53	R
1986639	59	R	52	R
2165962	11	R	<b>8</b>	R
2263628	66	R	42	R
2366891	57	F	54	F
2579944	51	R	47	R
2634050	46	F	46	F
3127931	43	F	53	F
3378554	51	R	30	R
3379024	43	F	54	F
3493911	43	R	45	R
3549196	44	F	40	F
3797824	41	F	49	F
3844682	64	R	54	R

F/R: forward or reverse, insertion direction,

**Bold:** count/avg. depth<10%, ND: not detected

表三十一、C00041 PacBio 分析挑選菌株抗藥圖譜

菌株 地理區	抗藥性	RIF	INH	EMB	SM	CAP	KM	AM	FQ	ETH	PZA
mCDC016	DST	R	R	R	R	S	S	S	S	R	S
花蓮縣秀林鄉	Illumina	<i>rpoB</i> S450L	<i>fabG1</i> C-15T, <i>katG</i> S315T	<i>embB</i> D1024N						<i>fabG1</i> C-15T	
mCDC104	DST	R	R	R	R	S	S	S	R	S	S
屏東縣泰武鄉	Illumina	<i>rpoB</i> S450L	<i>fabG1</i> C-15T, <i>katG</i> S315T	<i>embB</i> Q497P D1024N	<i>rrs</i> a906g				<i>gyrA</i> D94G	<i>fabG1</i> C-15T	
mCDC107	DST	R	R	R	S	S	S	S	R	R	S
花蓮縣吉安鄉	Illumina	<i>rpoB</i> S450L	<i>fabG1</i> C-15T, <i>katG</i> S315T	<i>embB</i> Q497P D1024N					<i>gyrA</i> D94G	<i>fabG1</i> C-15T	
mCDC108	DST	R	R	R	S	S	S	S	S	R	S
花蓮縣萬榮鄉	Illumina	<i>rpoB</i> S450L	<i>fabG1</i> C-15T, <i>katG</i> S315T	<i>embB</i> Q497P D1024N						<i>fabG1</i> C-15T	

表三十二、PacBio 定序 C00041 挑選菌株之基因體組裝結果

isolate no.	total output (Mb)	contigs # (HGAP)	contigs # (circulator)	genome size (bp)	max contig size (bp)
mCDC016	247.53	5	5	4,386,816	1,933,969
mCDC104	579.81	4	3	4,387,687	<b>4,374,503</b>
mCDC107	201.56	20	13	4,367,175	712,218
mCDC108	478.80	3	<b>1</b>	<b>4,389,983</b>	<b>4,389,983</b>

表三十三、mCDC108 Illumina 定序結果比對 PacBio 定序深度小於 40%區域

locus_tag	gene	bp	GC%
Rv0169	mce1A	3	66.67
Rv0746	PE_PGRS9	334	91.32
Rv0746	PE_PGRS9	225	89.78
Rv1068c	PE_PGRS20	281	85.41
Rv1091	PE_PGRS22	209	87.56
Rv2081c	(unknown func.)	28	57.14
Rv2741	PE_PGRS47	29	89.66
Rv3345c	PE_PGRS50	115	84.35
Rv3508-Rv3509c	PE_PGRS54-ilvX	140	81.43
Rv3512	PE_PGRS56	147	87.07
Rv3514	PE_PGRS57	291	85.57



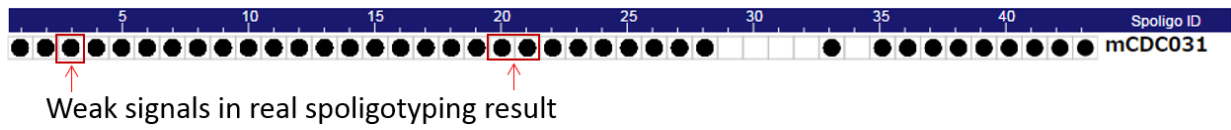
表三十四、DNA 萃取方法比較

WGS platform	Culture	MTBC inactivation	Cell lyse	DNase degradation	Cell debris removal	DNA precipitation	RNA removal	
Illumina	L-J or 7H10 agar medium	80°C 1h	Traditional phenol/chloroform method					RNase (final RNA<10%)
			lysozyme 37°C 2h	proteinase K 56°C overnight	CTAB/NaCl; chloroform, isoamyl alcohol, phenol	Isopropanol, ethanol		
PacBio			Commercial Kit (Gentra DNA purification kit, QIAGEN)					RNase (RNA Seq can be removed by analysis pipeline)
			Lytic Enzyme Solution 37°C 30 min	Cell Lysis solution 80°C 5 min	Protein Precipitation Solution	Isopropanol, ethanol		

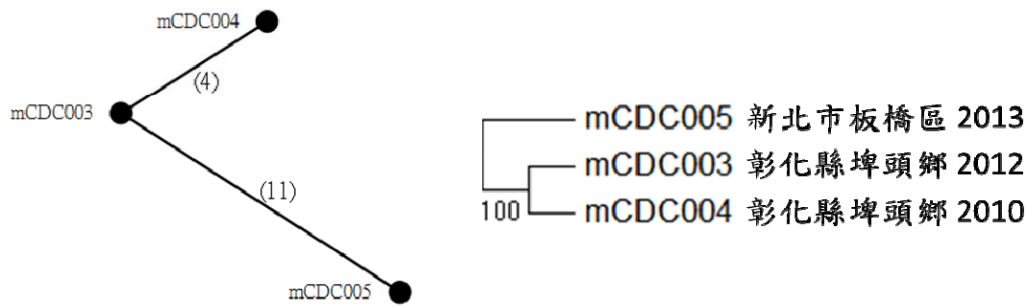
表三十五、結核菌 Illumina 全基因定序分析工具

TB WGS 分析套裝工具	抗藥性	基因型 (lineage)	演化樹	傳播鏈 (Network)	定序原始檔品質管理	優/缺點
JP/GenEpid-J /TGS-TB	v	v	v	v	v	單鍵上傳網頁工具。功能完整。AMR 抗藥預測位點經藥敏實驗驗證，並提供 TB profiler 結果參照。提供現行基因型方法學實驗預測。可完整下載比對分析檔核對定序比對結果。
UK/PathogenSeq /TB profiler	v	v	x	x	x	單鍵上傳網頁工具。抗藥性與基因型分析資料庫完善。
PhyResSE	v	v	v	x	v	單鍵上傳網頁工具。提供抗藥性、基因型與演化樹分析。
Mykrobe predictor	v	v	x	x	x	安裝版軟體，介面友善免上傳可離線使用。抗藥預測敏感性不足。

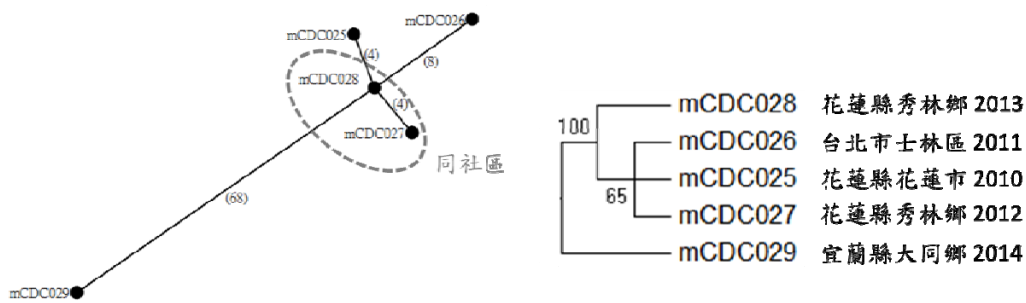
圖一、C00042 中的混合菌株 spoligotyping 預測與實驗結果比較



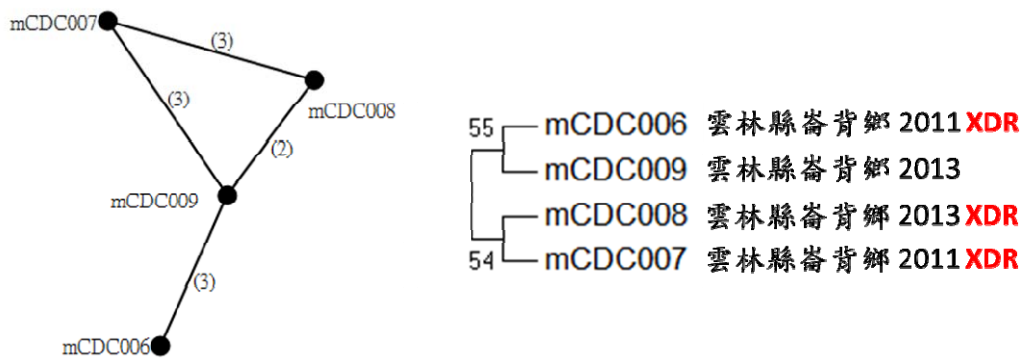
圖二、C00002 聚集菌株 SNV 差異數及演化樹分析



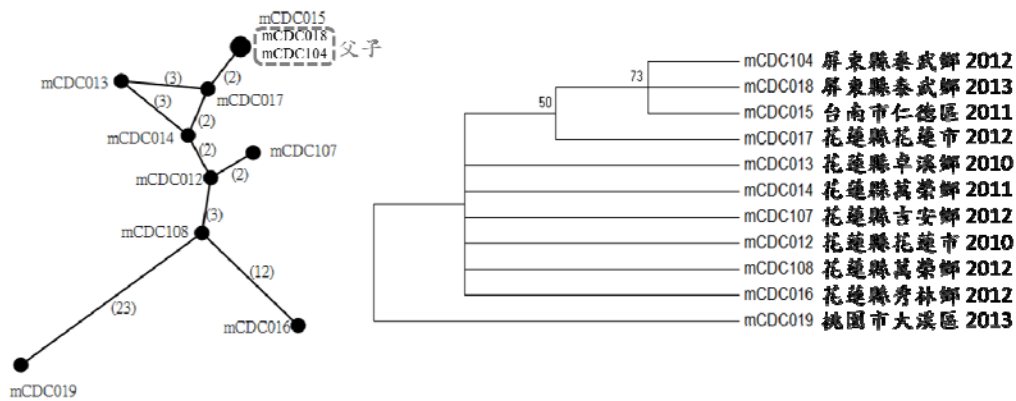
圖三、C00009 聚集菌株 SNV 差異數及演化樹分析



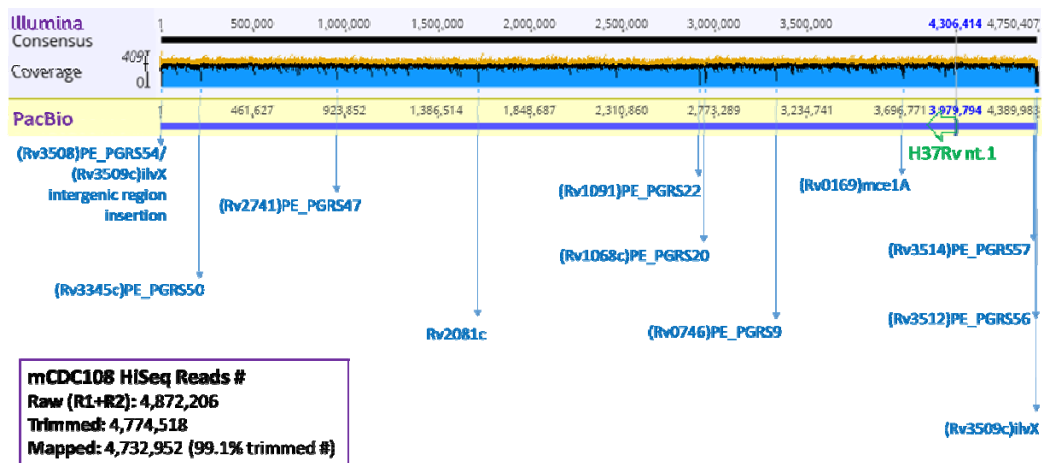
圖四、C00015 聚集菌株 SNV 差異數及演化樹分析



圖五、C00041 聚集菌株 SNV 差異數及演化樹分析



圖六、mCDC108 PacBio 比對 Illumina 結果



圖七、C00041 聚集菌株抗藥位點與演化樹對照圖

