

出國報告（出國類別：進修）

加州大學柏克萊分校  
勞倫斯柏克萊國家實驗室  
基因體學進修報告

服務機關：台中榮民總醫院

姓名職稱：劉伯瑜 醫師

派赴國家：美國

出國期間：August, 2014 ~ July, 2015

報告日期：104 09 07

## 摘要

感染症對許多國家而言，是嚴肅的國安議題。不論是新興傳染病或是抗生素的抗藥性，有效的研究與防治都需要大量高品質的基因體相關數據。

基因體的定序、序列組裝與功能註解，整道生產線都需要精細繁複的基礎建設，才能得到高品質的基因體數據。包括檢體的收集、基因庫的製備、定序等等，都需要嚴格的品管監控與不斷的流程改善，很多時候甚至需要修改原先廠商的設計以求得最佳化的實驗結果。對於定序後的數據，生物資訊部門的參與更是不可或缺。在許多基因體組裝與功能註解的策略上，常常需要不同部門的良好溝通才能產出最適合研究者所需的資料。

近年來，基因體學不僅以”讀取”為滿足，更進一步開始”寫入”。DNA合成技術讓我們能更精確的分析許多基因的功能，甚至在無法培養的微生物上。整合總體基因體、單細胞基因體，在目前將是發現自然界次級產物與其功能最有效的研發模式，進而重建整個代謝途徑。新一代基因編輯技術的技術，更能在活體細胞上進行基因編輯，未來發展不可限量。

## 目 次

出國報告提要.....	1
內容摘要.....	2
報告內文.....	4
建議事項.....	9

## 報告

### 目的

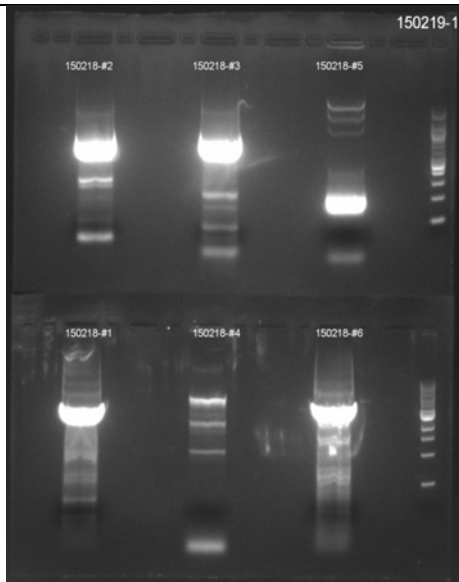
此次進修的主要目的為基因體學的最新發展與應用。進修地點是美國的勞倫斯柏克萊國家實驗室。勞倫斯柏克萊國家實驗室（Lawrence Berkeley National Laboratory）是美國聯邦政府設立的國家實驗室，由加州大學柏克萊分校代管。單位內有 13 位諾貝爾獎得主，2014 會計年度預算為 7 億五千萬美元，現有 3232 名員工。早期實驗室為發展粒子回旋加速器，及其後的核子武器計畫，需要大量科際整合而組成。也因此團隊合作一直是 Berkeley Laboratory 的傳統。在二戰期間主要的貢獻包括原子彈、雷達與近接信管的發明。戰後，Berkeley Laboratory 依然與軍方及產業維持良好的合作，並將觸角擴及更多研究領域。冷戰時期，為了分散風險，Berkeley Laboratory 開始轉型為以非機密 (Unclassified) 的研究為主。機密的研究則逐漸移轉至新墨西哥州的 Los Alamos National Laboratory (LANL) 與 Lawrence Livermore National Laboratory (LLNL)。期後 Berkeley Laboratory 在生物醫學相關的貢獻包括：人類基因體計畫，膽固醇的分類，氫氣對健康危害的發現，基因微矩陣晶片的發明，ECM 在乳癌的腳色，瘧疾感染紅血球的機轉，正子造影，以及微生物的基因改造。

在生物科技領域裡，目前 Berkeley Laboratory 整合了三大團隊，包括 Genomics Division、Life Sciences Division 與 Physical Biosciences Division。其中的 Genomics Division 就是這一年主要進修的單位。

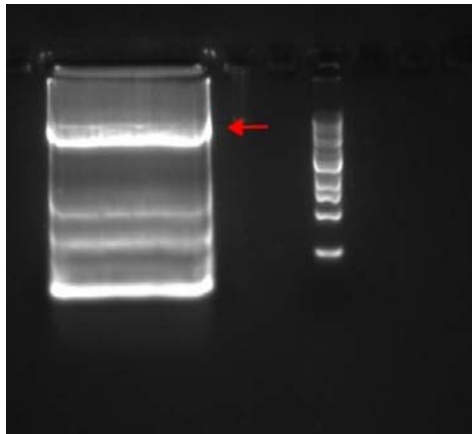
### 過程

前六個月在 Single cell genomic group 研修。學習製備定序所需的 DNA 基因庫，製作全基因體定序所需的品管操作與流程。Function Genomic Annotation 所需的生物資訊軟體操作。

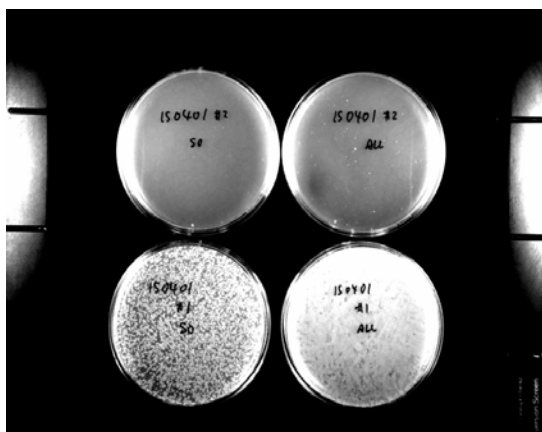
後六個月在 Synthetic biology group 研修，學習基因編輯技術。



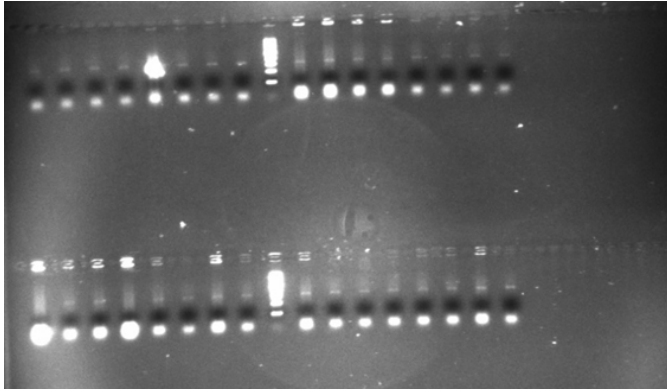
用引子從各種質體得到 DNA



從膠體萃取後用 組裝成一個選殖載體



基因轉殖後挑選細胞株



用 PCR 篩選可能轉殖成功的細胞株後送定序

## 心得

### 感染症與國家安全

這一年來，身處聯邦的生物實驗室，從日常作息裡的會議、公文、郵件、甚至同事的談論裡，深刻的體會到美國是不折不扣地將感染症當作一個國安議題在操作。舉微生物抗藥性為例。歐巴馬政府在 2014 年 9 月宣布，將提撥超過 12 億美金的預算來解決抗藥性的問題。期間將整合聯邦各級部門，包括醫療、公衛、獸醫、農業畜牧、食安、研發等機構。舉全國之力，期望在 2020 年達到減少抗藥性的威脅。其行動計畫中特別要求需要大量高品質的微生物基因體資料，作為分子流行病學及防疫的重要參考。而部分原本以研發戰略性武器為主的先進實驗室，近年來也陸續增設 Bioscience 部門。

以新興傳染病而言，美國國防部早在 1995 年就已經完成伊波拉病毒在靈長類傳染途徑與致死劑量的研究。爾後開始不遺餘力地與藥廠合作進行新藥研發。2014 年的西非伊波拉病毒大流行方能藥物得以使用。全基因體定序是新藥研發的必須工具。為了更有效的監測疫情，美國陸軍甚至直接在賴比瑞亞建立了一個基因體定序中心。

不僅如此，對於微生物相關的研究，聯邦相關單位一直都在進行深入的研究。舉例來說，國家情報總監 (Director of National Intelligence) 下轄的情報高等研究計劃署 (Intelligence Advanced Research Projects Activity) 對於病毒、細菌及相關的毒素基因體有著深入的研究，並且進一步探討人體在曝觸這些物質之後的反應，其中一個目標為發展能在曝觸病原後 30 分

鐘內能夠診斷的檢測方法。國防部下轄的國防高等研究計劃署(Defense Advanced Research Projects Agency) 更在 2014 年成立生物科技部門，工作重點之一是運用大量的基因體數據，診斷與治療各式新興與再浮現傳染病上。

### 基因體學與感染症

從上述內容不難發現，許多傳染病的研究都需要基因體資料。而基因體學與感染症的關聯何在？簡單的說，基因體學的發展將從根本上改變臨床感染症的全貌。將是完全的典範轉移 (Paradigm Shift)。例如，過去我們對微生物的鑑定，大多是用生化反應的表現來進行種的分類。而近十年來大量的微生物基因體數據不僅將改變我們的鑑定流程，更可能在不久的未來完全改變原核生物的分類方法與準則。以 Berkeley Laboratory Genomics Division 最近的發展為例：實驗室發展了一套稱為 Microbial Species Identifier (MiSI) 的方法進行菌種鑑定。它運用基因體定序資料中的 genome-wide Average Nucleotide Identity (gANI) 以及 alignment fraction (AF) 分析兩組基因體的親緣與演化樹關聯，進一步進行菌種的歸類。Neha Varghese 在發展這套演算法時，大量使用了位於 Oakland 的 National Energy Research Scientific Computing Center 裡的超級電腦。他們的成果發表在今年 7 月 Nucleic Acids Research 的文章裡。

Berkeley Laboratory 更進一步運用這些技術，改寫我們對生命體的認識。過去 40 年來，我們將地球所有的細胞生命體(cellular life)分類為 3 個領域：真核生物(Eukaryota)、細菌(Bacteria) 與古菌 (Archaea)。這幾年，隨著基因技術的進步，實驗室對於大量無法培養的微生物基因體進行演化樹分析後發現第 4 種細胞領域的存在的可能性。在幾個月前發表在 Science，由 Berkeley Laboratory Genomics Division 微生物計畫主管 Tanja Woyke 以及 Genomic division 負責人 Edward M. Rubin 聯名發表的文章裡，進一步詳加描述如何應用現有的基因體技術，發現地球上新生命形態的計畫。

### 基因體定序流程

Berkeley Laboratory Genomics Division 現階段以次世代定序 (next-generation sequencing, NGS) 為主力。現有 Illumina MiSeq、Illumina HiSeq、以及 PacBio RS。

其流程大致為將基因體片段化，建立基因庫，以聚合酶連鎖反應 (PCR) 擴充後做定序和

資料分析。(也有不做 PCR 直接定序的 protocol) 將基因體片段化的過程，已有部分開始自動化，但是這裡大多的實驗室還是手動操作。流程包括 Sonification 以及一系列的 Enzymatic Reactions-- End repair, A-tailing, adapter ligation。在這部分花了許多時間練習才製作品質好的 library，通過定序前的品管試驗。

在整個學習的過程裡，最深刻的體會，不僅是龐大經費的投入。更在於 Berkeley Laboratory 的工作人員，能夠花費大量的時間與精力進行整個流程的改善。雖然這部分的努力不容易被外界發現，也不容易產出高影響力的文獻。但是這一份基礎建設的完成，卻是所有後續研究不可或缺的核心。對所有基因組的研究，更有深刻的影響。

雖然這幾年，桌上型次世代定序機越來越廉價，也越來越多研究單位與公司投入這一個市場。然而，產出數據的品質良窳差異非常大。這也是來到 Berkeley Laboratory 才認識的一點。基因體數據的產出、組裝、註解充滿許許多多可能出錯的地方。沒有許多人不斷的努力在改善流程，產出的數據勢必出現許多錯誤。

近幾年有越來越多的文獻指出，現有許多發表的基因體數據充滿各種錯誤。例如，一篇分析 2749 份非靈長類基因體資料的研究顯示，492 份序列中出現汙染的人類 DNA 序列。這些研究在在顯示了一個好的定序中心裡，品管與不斷追求流程最佳化的重要。甚至很多時候都需要修改原先廠商的設計以求得最佳化的實驗結果。例如，用在篩選 DNA 大小的 AMPure XP Beads，Berkeley Laboratory 做了不少實驗累積資料，改進後的實驗步驟可以比廠商提供的實驗步驟更精準的篩選出各種大小的 DNA。又如 Illumina sequencer 原本建議的基因庫大小為 200~300 bp。Berkeley Laboratory 改良後可以做出超過 700 bp 的基因庫。對於後續的基因體組裝有很大的幫助。Berkeley Laboratory Genomics Division 在 2014 年可以產出 1058460 億個 bp 的定序資料，靠的不僅是經費，還有不斷改善的基礎建設與生產線。

## DNA 合成

隨著大量基因體數據的產生，衍生出越來越多的未知序列。特別是總體基因體學的發展發現許多無法在自然界分離的 DNA 分子。DNA 合成的技術讓我們能將定序資料結合基因與代謝途徑，合成 DNA 分子，再轉殖入微生物進行功能分析。2012 年已經有 2 百萬個 bp 的 DNA 在 Berkeley Laboratory Genomics Division 被合成。部門主管計畫在 10 年內能夠完



成整個基因體規模的合成作業。加上近年來日新月異的基因編輯技術，讓我們有機會在活體內進行基因編輯。未來的應用無可限量。

### **建議事項**

1. 未來的生技產業，需要大量的高品質基因體資料。能產出好的基因體資料，需要集中資源，強化基礎建設才能達成。即使在美國，能符合此一標準的定序中心也不多。70年代若無工業技術研究院，任憑各研發單位各自發展，今日台灣當是另一種面貌。國內應審慎思考整體生技產業的戰略，建立以 User Facility 概念的基因體中心。
2. 在 2015 年 6 月的 Nature 雜誌裡，將基因編輯技術比擬為自 PCR 發明後最具破壞力的生物科技發現；Science 雜誌的專刊則以革命形容這項技術。在過去半年，多家國際生技公司與藥廠紛紛投入此一領域並開始專利佈局。國內宜審慎評估，尋找台灣能切入該市場的利基。
3. 以研究能量來說，一年期的訓練時間對於當今的尖端科技而言，委實時間過短。特別是在競爭激烈的學術單位，許多創新的研究發現與應用，必須建立在長期的基礎建設、數據累積與科際整合上。需要較長時間的投入該組織，方能窺其全豹。進而對國內科技發展有所助益。